

Embargoes on science?

Researchers should know more than has been disclosed about the symbiotic relationship between the places where they publish and the general press.

THE *New England Journal of Medicine*, with *The Lancet* one of the most effective research journals as well as a means of keeping physicians in touch with medical research, seems have stumbled into an uncharacteristically undignified quarrel with an even more successful communications agency — the Reuters news agency. The row is entertaining: public rows usually bring out the unexpected in the participants. But some of the issues raised are also instructive for the scientific community as a whole.

The circumstances are these. The journal (henceforth *NEJM*) distributes copies of each weekly issue to selected journalists and periodicals on the written understanding that no reference is made to the contents until the formal date of publication, by which time the issue is safely in the mails and in the hands of at least some of its subscribers. In January this year, a Reuters journalist put on his agency's wires a news item about a research study suggesting that ordinary aspirin taken every other day can help people to avoid heart attacks. Sadly for the journalist, his agency and everybody concerned, the authentic account was due to appear in that week's *NEJM*. Reuters was told that the privilege of its advance copy would be withdrawn for six months, and was asked to give a more solemn form of the usual undertaking before the privilege could be restored. Reuters, saying that its reporter had learned what he wrote about aspirin and heart attack without seeing the relevant copy of the *NEJM*, declined to give the undertaking asked for. It will presumably now be medically less well-read.

Readers may be surprised to know about these advance copies, but *NEJM* is not the only journal that follows this practice. So, too, does *Nature*. Most general journals that offer their contributors wide readership follow some practice of this kind, always on the understanding that no use will be made of advance material before publication. One objective, often quoted, is that it gives journalists much-needed time in which to prepare stories intended for general consumption. Another, less often acknowledged, is that the practice puts all interested journalists on an equal footing, so that individual reporters in science or medicine are freed from the need to determine what they write about by the knowledge of what their rivals have already written, which gives the journals the benefit that their contents are likely to be noticed in writing by more periodicals than would otherwise have done so. Evidently the breaking of the embargo undermines this mutually beneficial arrangement.

Researchers should know that these practices constrain them

as well as reporters; contributors to *Nature*, for example, are told when their articles are accepted for publication that it is a pre-condition that there should be no pre-publicity for their work. It is also common knowledge that many journals reject that approach. The American Physical Society, for example, on behalf of its several excellent journals, has taken the line that contributors can seek whatever advance publicity they choose for their discoveries, largely on the grounds that research which is mostly paid for by taxpayers should be freely accessible as soon as possible. The American Physical Society has even offered to help those looking for publicity find a way of doing so.

Each of these positions is, of course, tenable, but on a different ground. One calculation that stands out is that the journals insisting that pre-publicity by their contributors is against the rules must of course be sensibly flexible about their interpretation of what the rules imply. People who talk at meetings at which journalists happen to be present cannot (and should not) avoid spilling the beans; journalists smart enough to recognize what is being said are entirely within their rights to make what use they wish of what they learn. The embargo on pre-publicity becomes a kind of formality. It is a different matter when a contributor supplements what he has to say at a meeting by thrusting copies of his paper into the hands of journalists or when he sends the essence of his latest contribution to a reporter at one journal before he has had the time to send the whole of it to the journal in which he would like to see it preserved for the rest of time (one of *Nature's* recent trials) — or when a contributor's university press department chooses to put out a statement on the subject in advance of publication.

Most journals are flexible and forgiving in their way of dealing with the issues that inevitably arise in operating such a policy. How can it be otherwise, when one of their interests is their own advantage, not just the general well-being? Fairness also becomes an issue. There are indeed occasions when enterprising journalists are able to steal a march on the more specialist press, which is generally to be welcomed. Indeed, the journals that look askance at newspapers which beat them in speed of publication had better quickly recognize that there is a kind of inbuilt logic in the diurnal rhythm that they cannot loftily ignore.

It must also be confessed that there are also occasions when it seems that the gun has been jumped implausibly. Some of the tales that are told are even on the unbelievable side of the implausible. People will sometimes say that it is not their fault that an account of an article appearing in, say *Nature* appeared before it should have done because of an oversight by an editor. (More commonly, people explain that it was not they who forget to mention the source of their inspiration, but that of the person who sent their copy to the printer.) But these excuses are neither common nor persistent. Most journalists seem aware of the benefits of the system. So, mostly, do the journals. Maybe, on this occasion, the *NEJM* has been a little heavy-handed, at least in its assertion that its chief concern is to avoid the situation that may arise when patients learn of new treatments from a news article before the their physicians have had a chance to read the authentic version. Are physicians that diligent, or the mails that good? □

NEWS, pages 192–198

Vitamin-D patent legislation ■ Stanford ■ Blood plasma in Japan ■ Ariane launch ■ AIDS network ■ Surrogate cows ■ Postgraduate medical research ■ International science education ■ AIDS agreement ■ High-tech city for Australia? ■ Landsat 6 ■ UK university earnings ■ ESA's Olympus ■ Neuroscience collaboration ■ South Africa ■ Indo-US vaccine programme ■ West German bacterial release ■ UK academic science ■ CNRS/industry links



Legislation on vitamin D patent makes waves

- Lawyers may be the beneficiaries
- Wisconsin U. pinch-hits for courts

London

A CASE heard in the Chancery Division of the British High Court last Tuesday (8 March) has drawn attention to a long-standing battle over patents for processes to manufacture the active principle of vitamin D, 1 α ,25-dihydroxycholecalciferol, but observers differ in their interpretation of the judgement by Mr Justice Whitford, which has not yet been made public.

The patent dispute is significant not merely for the costs which have accumulated over almost fifteen years and the importance of a means of synthesizing the active principle of vitamin D (formed in the liver and kidneys by the successive conversion of the natural vitamin), but for the distinction of those responsible for the two alternative processes — Sir Derek Barton, now at the University of Texas at Austin, who was at Imperial College London when he won the Nobel Prize for chemistry in 1969, and Dr Hector DeLuca, professor of biochemistry at the

University of Wisconsin.

The litigants in the long-standing suit are not the inventors of the two processes, but the organizations to which the inventors assigned the patent rights. The Wisconsin patent (based on E.J. Semmler, M.F. Holick, H.K. Schnoes and H.F. DeLuca, *Tetrahedron Lett.* **40**, 4147; 1972) has been assigned to the Wisconsin Alumni Research Foundation (WARF), originally constituted in 1925 by Dr Harry Steenbock, who was among the first to demonstrate the value of vitamin D in the treatment of rickets. Last year, WARF contributed more than \$11 million to the university's research fund.

The Barton patent (D.H.R. Barton, R.H. Hesse, M.M. Pechet & E. Rizzardo *J. Am. chem. Soc.* **94**, 9518; 1973) is assigned to the Boston-based company Research Institute of Medicine and Chemistry (RIMAC), of which Pechet is the chairman. The US patent applications were lodged in the early 1970s, since when the two holders were in litigation until their dispute was settled out of court last summer.

The terms of the agreement were 'sealed' by the US courts, but include an agreement that the litigants will not in future attack each others' patent rights to the active principle of vitamin-D and the payment of an undisclosed sum by WARF to RIMAC.

Quite apart from the duration of the litigation, the way in which it has been punctuated by personal charges against the Wisconsin group has made the case unusually bruising. DeLuca said on the telephone last week from Wisconsin that he had been so resentful of some of the charges that he had asked the University of Wisconsin early last year to institute a formal inquiry by outside people.

Dr Bernard C. Cohen, vice-chancellor for academic affairs at the university, said on Monday this week that DeLuca's request last spring had been delayed because the university had been hoping that the suit then current in the US courts would clear the air. The out-of-court settlement meant that the issues did not come to trial.

During the past few weeks, Cohen said, he has learned that recruiting a panel of people sufficiently knowledgeable about laboratory work and the field concerned, but sufficiently detached from the participants "is not as easy as I had hoped, but as difficult as I had feared".

Part of the contention in the case stems from the detail in which the discovery process built into US patent law has enabled the RIMAC side to inspect the Wisconsin laboratory notebooks, from which it appears to have emerged that the Wisconsin synthesis as published would not have worked as described.

Dr M.F. Holick, a graduate student in DeLuca's laboratory at the time and a participant in the synthesis on which the WARF patent was based, acknowledged on Monday the truth of one of the charges made in litigation by RIMAC that one of the fourteen stages in the process, the purification of a pair of isomers (only one of which was needed for the later steps), would normally yield only the unwanted isomer, because of the unexpected difference of the diffusion constants of the two isomers.

Holick also confirmed this week a point made by RIMAC in the litigation that the published synthesis is now known to have been successful because one of his colleagues, Dr Eric J. Semmler, having broken a flask and spilled the output from the chromatography column on the floor, returned the contents to the top of the column for repurification, whereupon the two isomers eluted together.

It has also been an important issue in the litigation that DeLuca first learned of the Barton synthesis when asked to act as a referee for its first publication in the communications section of the *Journal of the American Chemical Society*, and that the Barton process was used as the means of making small quantities of the active principle of vitamin D available for tests of biological efficacy.

DeLuca said on the telephone last week that he had given the Barton manuscript a favourable and prompt report (contrary to suggestions during the trial) and that it was in no sense improper to use another's published processes for the purposes of research, which is of course generally accepted.

Despite the out-of-court settlement between RIMAC and WARF, litigation nevertheless continues. Last week's hearing in London was an appeal by WARF against an order by a British patents examiner last November that the files in the case should remain open, allowing the British Patent Office to continue brooding about the validity of the original WARF patent. WARF may yet appeal.

But one set of documents has been returned to WARF's representatives. WARF seems also to have won the point that the amendment of its patent does not, of itself, impugn its validity.

Meanwhile, RIMAC, according to its London lawyers, is distressed that WARF is disputing the RIMAC patent in New Zealand (where WARF has no patent) despite last year's settlement in the United States.

John Maddox

Stanford still on top

Berkeley

Stanford University has received over \$6 million in royalty income from patents in 1986-87, up \$1 million from the previous year. The sum was more than that received by Harvard, the Massachusetts Institute of Technology (MIT) or the University of California (UC), and possibly higher than any other US university, according to Kathy Ku of Stanford's technology licensing office.

For the first time, Stanford's leading earner was the 1980 patent by Stanley Cohen of Stanford and Herbert Boyer of the University of California at San Francisco on recombinant DNA.

Stanford shared the royalty income with UC. The Cohen-Boyer patent brought \$1.7 million to Stanford, surpassing the former leader, a 1971 patent on the FM synthesizer chip, invented by music professor John Chowning, and used in Yamaha music synthesizers.

Included in last year's income total was \$700,000 from the settlement of a patent infringement lawsuit with Coulter Corporation of Hialeah, Florida. In addition to the \$700,000, Coulter will pay royalties for its use of phycobiliproteins, patented jointly by Stanford and UC as a fluorescent label for cell sorting.

Marcia Barinaga

Pharmaceutical companies profit from Japan's blood boom

Tokyo

PROFITEERING by pharmaceutical companies and hospitals at the expense of patients and the national health insurance system may be behind Japan's skyrocketing consumption of blood plasma, according to an investigation by Japan's *Mainichi* newspaper.

Japan has already been severely criticized by the World Health Organization for its excessive blood imports, which monopolize about a third of the world market (*Nature* 317, 101; 1987). Japan's consumption of blood plasma soared to nearly 4 million litres per year in 1985, nearly all of it imported from the United States. *Per capita* consumption of blood plasma is about three times that of Britain (see figure).

The Ministry of Health and Welfare has tried to curb use of imports by doubling the volume of blood allowed to be taken from Japanese donors and by issuing guidelines to try to prevent excessive prescription of blood products by the medical profession. But there has been only a slight drop in consumption.

A number of factors lie behind the import boom. Cheap imports of concentrated blood coagulants for the treatment of haemophiliacs came on the market in 1978 and quickly replaced cryoprecipitate which had to be made from fresh domestic blood. The new coagulants could be self-administered and sales soared as haemophiliacs began storing them in their refrigerators at home.

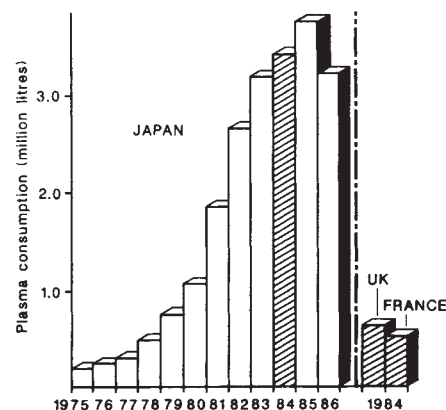
Even when it became apparent in the United States in 1983 that AIDS could spread through the use of these coagulants and the United States and Europe turned to heat-treated products, Japan's imports of the untreated coagulants continued to rise, reaching a peak in 1985 when heat-treated products were finally allowed on the market. Meanwhile, many of Japan's haemophiliacs are believed to have been infected with AIDS (*Nature* 331, 552; 1988).

Also behind the blood boom has been increased use of albumin as a 'nutrient' for Japan's rapidly increasing population of elderly people. But the prime driving force behind the rise in use of blood products, according to the *Mainichi*, has been enormous discounts offered by the pharmaceutical companies.

The Ministry of Health and Welfare sets 'official' prices for all drugs sold in Japan. But, according to the *Mainichi*, the pharmaceutical companies have been selling blood products to hospitals at 35–85 per cent below the official price. As the prescription charges for patients and the national health insurance system are

based on the official price, the hospitals have been able to reap enormous profits by prescribing more and more blood products, the newspaper charges.

On Monday, a ministry official would neither deny or confirm the *Mainichi* report. He said that it was "under study". But the official did say that the prices charged by the pharmaceutical companies are a matter for them to decide.



Japan's consumption of blood plasma.

How could the companies offer such large discounts? According to the *Mainichi*, in 1985 non-heat-treated coagulants could be bought in bulk in the United States for as little as a fifth of the official price in Japan and in that year imports reached a peak of nearly 160,000 bottles.

Clearly the "official" prices of blood products in Japan need to be re-examined and prescription of blood products strictly controlled. The dangers of AIDS infection have been greatly reduced through blood screening and the use of heat-treated coagulants, but infection is still possible with hepatitis non-A non-B which is not killed by heat treatment.

The national health insurance system is already in trouble, with income failing to cover expenditure, so the incentive for the ministry to slash excessive use of expensive blood products is strong. But, instead, the ministry's policy is to increase domestic production of blood plasma.

The ministry hopes to secure one million litres of plasma by 1990 through increased blood donations. But already over 8 million people, seven per cent of Japan's population, donate blood each year, a very high rate of donation.

So why does the ministry not simply slash Japan's blood plasma consumption to internationally acceptable levels? The ministry's scientific committees are manned by medical doctors who often collaborate with pharmaceutical companies in research. It is these doctors and the pharmaceutical companies who really determine policy.

David Swinbanks

Launch of Ariane boosts flagging telecommunications

Paris

DURING the night of 11 March, Arianespace, Europe's commercial space consortium, successfully put French and US telecommunications satellites into orbit from an Ariane-3 rocket.

The launch, the twenty-first since the Ariane programme was set up by the European Space Agency in 1973, will bring much-needed relief on both sides of the Atlantic. While French telecommunications has been beleaguered since the Telecom 1B satellite broke down on 15 January (see *Nature* 331, 469; 1988), the grounding of the Challenger shuttle has meant that no US commercial satellites have been launched in two years.

Telecom 1C, the French satellite on board Ariane flight V21, will replace the failed Telecom 1B and will take the strain off the ageing Telecom 1A, which has to take the extra traffic as an interim measure. For the past two months, Telecom 1A has been relaying all French radio and television broadcasts, military communications and overseas telephone calls — a task which it could not continue to handle without shortening its life.

Engineers at Matra, the French manufacturers of Telecom 1C, will still be biting their nails, however. The satellite is essentially a duplicate of its failed predecessor, although it was sent back from the Arianespace launch site in French Guiana for thorough testing in February, delaying the V21 launch. Last month the West German telecommunications department had to abandon its brand-new direct-broadcast television satellite, TV-SAT, which failed to come on stream despite a faultless Arianespace launch.

The other satellite put into orbit on flight V21, Spacenet IIIR/Geostar, belongs to the US telecommunications company, GTE Spacenet. Geostar is a radiolocalization satellite, immediately destined for use by road and rail transport companies, but likely to be extended for use by shipping. Since the mid-flight explosion of the US National Space and Aeronautics Administration (NASA) Challenger shuttle in 1986, only military payloads have been put into space on US rockets. NASA now hopes to resume scientific and commercial satellite launches this summer (see *Nature* 330, 195; 1987).

Having got back into its stride after the setbacks of the past two years, Arianespace intends to keep up the pace with 8 launches scheduled for 1988 and 9 each year until 1990. The company has a backlog of 43 satellites, worth over \$2,600 million.

Peter Coles

New strategy sought for UK postgraduate medical research

London

A SWEEPING review of Britain's largest body for postgraduate medicine is under way, aimed at developing a coordinated policy for research and teaching, moving away from the traditional discipline-based system.

In the face of declining support from central government and increasing reliance on short-term grants from charities and industry, the British Postgraduate Medical Federation is hoping to devise a research infrastructure that will more efficiently harness the federation's scientific expertise and attract longer-term investment from non-government sources.

The federation is a school of the University of London comprising eight postgraduate institutes — Hunterian (the scientific basis of surgery), child health, cancer research, dental surgery, neurology, ophthalmology and psychiatry — and the departments of the postgraduate deans in medicine and dentistry of four public health regions.

Last spring, the federation's new president, Professor Michael Peckham, set up a steering group to consider ways of improving scientific coordination between the institutes, which between them have a total of 92 academic departments, almost 2,000 members of academic staff and 7,500 postgraduate students.

A first draft of the federation's plan was

presented to the university authorities last week. The plan shows how government support through the University Grants Committee (UGC) is accounting for an increasingly small proportion of the federation's total income.

In 1980–81, recurrent grant was £6.6 million, rising by 16.9 per cent to £7.7 million by 1986–87. This contrasts with an increase in income from non-UGC sources over the same period of 139.7 per cent, to £11.98 million.

Between 1980 and 1987, 132 posts were frozen or lost at the federation, with 152 new posts being gained. However, excluding those at the Institute of Cancer Research, of the posts lost, 97 per cent were UGC-funded, whereas of the new posts, only 15 per cent were UGC-supported. More than 30 per cent of the new posts were supported by charitable bodies.

The plan says that the short-term nature of many of the non-UGC positions "is likely to have serious implications in the long-term, particularly with regard to the career structure of non-clinical scientists". Peckham says that because fundamental scientific research in areas such as molecular and cell biology cut across traditional disciplines, increased cooperation and coordination between the institutes will result in stronger science. Peckham hopes that a more coordinated and

flexible research strategy will also lay the foundations for persuading non-government bodies to invest in longer-term support, with a consequent improvement in the career structures for the researchers. "The reduced attractiveness of medical research as a career option is causing considerable anxiety at the present time", Peckham says.

Simon Hadlington

AIDS workers going for computer link

Washington

THE World Immunological Network Project Foundation — an organization which promotes hospital videoconferencing — has ambitious plans to link all AIDS researchers worldwide together in computer networks, according to Genevieve Clavreul, the foundation's director.

Clavreul scored her first big success last week by persuading Robert Gallo of the US National Cancer Institute and Luc Montagnier of the Institut Pasteur in France to activate accounts with BITNET — a widely used scientific network in the United States — and the European Academic and Research Network (EARN). Members of Gallo's and Montagnier's laboratories will now be able to communicate electronically, through an interface between the two networks.

By tying together electronically two of the leading centres of AIDS research Clavreul hopes to stimulate increased use of data-sharing and communication networks among the biomedical scientific community. Her ultimate plans are to create a 'subset system' on AIDS research that would allow the exchange of electronic mail, and provide access to AIDS-specific databases and software. Clavreul is working with IBM to encourage biological researchers to tap into electronic mail networks. Most researchers studying AIDS work at institutions already have accounts on major networks such as BITNET and EARN, but biologists have been slower than other scientists to communicate electronically because their work involves less computing.

But despite Clavreul's enthusiasm, the idea of a specialized computer network for the exchange of AIDS information among researchers is not new. The US Presidential Commission on the HIV Epidemic recommended the creation of a computer network for AIDS research in its preliminary report (see *Nature* 332, 3; 1988). Recently, the American Foundation for AIDS Research convened a meeting of AIDS researchers to discuss whether or not such a network would be useful. The consensus was that communication by facsimile machine and telephone was filling current needs, and that frequent meetings were more beneficial.

Carol Ezzell

Exotic cows born from Indian surrogates

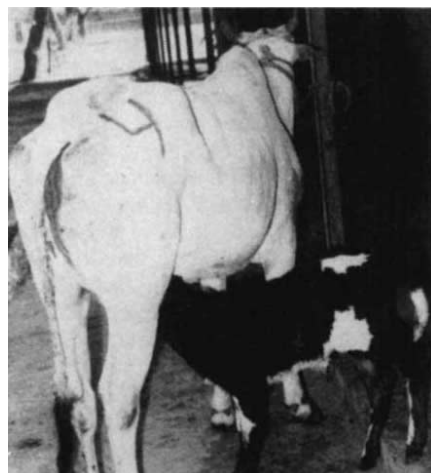
New Delhi

A STRAY Indian cow last week delivered a pure Holstein-Friesian (HF) female calf in what is said to be the first successful attempt to get exotic breeds using indigenous cows as surrogate mothers.

The birth was the result of non-surgical embryo transfer (ET) carried out by scientists of the National Institute of Immunology (NII) and the Indian Agricultural Research Institute (IARI) in New Delhi. The embryo used was from pure HF animals; the recipient was a nondescript cow of poor genetic quality.

India has some 85 million stray cows that yield little milk but are objects of worship by the Hindus. NII director Dr G.P. Talwar says the present work has demonstrated the feasibility of using these animals to create a stock of exotic breeds by making them pregnant with embryos from 'elite' animals. He says that one elite cow can, in principle, supply enough embryos a year to make 12 stray cows pregnant by super-ovulating the donor and flushing out the embryos at three-monthly intervals.

Although ET has successfully been



employed for some time in India, only crossbred cows have previously been used as foster mothers. Several stray cows in Haryana, adjoining Delhi, are pregnant by this method and it is expected that ET will gradually take over from artificial insemination as a standard procedure for genetic improvement of Indian cattle.

K. S. Jayaraman

International science education gets bad grades

Washington

A NEW study of school science achievement in 17 countries appears to predict disaster for the future of science in the United States, mediocrity for the Japanese and great success for Britain. But appearances can be deceptive: differences in educational systems and in the proportions of students who remain in school through to the final year greatly influence the final rankings.

The study, conducted by the International Association for the Evaluation of Science Achievement in 1983–86, reports the results of tests in the basic concepts of biology, physics and chemistry taken by some 44,000 students in their last year of school. All the students who participated in the study were enrolled in science courses. The results, showing Britain near the top of the rankings and the United States near the bottom, are summarized in the table below.

Specialization partly explains the successes scored in Britain, Hong Kong and Singapore. In their final years, students in these countries are concentrating their efforts in only two or three subjects. But more important is that only one-fifth of the age-group eligible to be in the last year of school is actually enrolled, leading the authors of the report to ask whether the high performance of this elite is obtained at the expense of mass education. In comparison, the United States retains three-quarters of its high-school-age children in school to the final year, and Japan retains almost all its students.

Even though the United States and Japan have good records for educating the majority of their populations, problems are still evident. Advanced science stu-

dents in the United States — those who have studied a subject for more than one year — perform poorly compared to the best students from Britain. The study concludes that “for a technologically advanced country, it would appear that a reexamination of how science is presented and studied is required”. Japanese science students show a large gap in their knowledge of biology compared to their performance in physics and chemistry, and significant differences in science achievement show up between students of various schools, suggesting that educational opportunity is not equitably distributed.

The difference in the level of science achievement between boys and girls is still high in most countries, according to the study. Boys performed better than girls in all subjects except biology in nearly every country; only in Australia, Hong Kong

and Sweden did girls score slightly higher than boys.

Three more reports will be issued in 1989 on how variations in science curricula affect performance, on specific differences in achievement between and within countries, and on changes in science achievement analysed by comparing test results from 1970 and 1983–86.

Carol Ezzell

AIDS agreement

London

THE principal medical research organizations of France and Britain have signed a formal agreement on cooperation in AIDS research. Under the agreement, the Institut National de la Santé et de la Recherche Médicale and the Medical Research Council will retain their own programmes of AIDS research but will take part in a variety of joint activities, such as meetings and exchanges. In addition, there may be joint financial support for some research undertaken in each country. Simon Hadlington

Japan's high-tech planners dream of Australia's beaches

Sydney

AUSTRALIA and Japan are to conduct a joint A\$5 million study of the feasibility of building an international high-technology city somewhere in Australia. The city, with a population of 250,000, would be built around government and private research institutes and high-technology industries. A total investment of some A\$1,000 million would be expected, most of it coming from private industry.

The plan is the brainchild of Japan's Ministry of International Trade and Industry (MITI), and has its origins in Japan's 'technopolises', regional centres chosen for government grants to help stimulate the growth of high-technology industry. But the Australian project is much bigger in scale. MITI has already coined a new word to describe the new city — 'multi-function-polis' or MFP.

MITI first put forward the plan early last year. A MITI delegation is now touring Australian state governments, all of which, with the exception of Tasmania, are clamouring for MFP to be built within their borders.

Given Japan's wish for the city to have a resort-like environment, the prime contender must be on Queensland's Gold Coast, near Brisbane. MITI wants the new city to provide researchers and their families with a lifestyle, integrating work and leisure, that cannot be achieved in overcrowded Japan.

Australia can offer Japan a good climate and plenty of space for the city. Brisbane is in a neighbouring time zone to

Tokyo. And Australia is seen as an advanced country with a strong research base which is not a serious competitor in high technology. Australia could also provide a neutral ground for researchers from Japan, Europe and the United States



to mingle. Research institutes and corporations from around the world, and particularly from the United States, will be invited to set up facilities in the city.

For Japan, the city would be an experiment in twenty-first century living which could become a prototype for other cities on the Pacific rim. For Australia, according to Senator John Button, Minister for Industry, Technology and Commerce, the city would be an opportunity for Australian industry to leap into a new realm of innovation and development.

Charles Morgan

Ranking according to science achievement in final year in school

Biology	Chemistry	Physics
Singapore	Hong Kong	Hong Kong
England	England	England
Hungary	Singapore	Hungary
Poland	Japan	Japan
Hong Kong	Hungary	Singapore
Norway	Australia	Norway
Finland	Poland	Poland
Sweden	Norway	Australia
Australia	Sweden	United States
Japan	Italy	Sweden
Canada	United States	Canada
Italy	Canada	Finland
United States	Finland	Italy

Data for Korea, the Netherlands, the Philippines and Thailand were not available for these categories. All data from *Science Achievement in Seventeen Countries* (US National Science Foundation, Office of Studies and Program Assessment, Washington, DC, 1988).

Prospects looking up for new remote-sensing satellite

Washington

AFTER months of delays and political wrangling, plans for Landsat 6, the new remote-sensing satellite destined to replace the ageing Landsats 4 and 5, seem ready to move ahead. Late last month, EOSAT, the commercial entity established by the government to operate the Landsats, and the National Oceanic and Atmospheric Administration (NOAA) reached an agreement that will permit Landsat 6 to be launched aboard an Air Force Titan II rocket by spring of 1991.

A year ago, the next generation of Landsats seemed to be going nowhere (see *Nature* 325, 187; 1987). Congress and the Reagan administration were at loggerheads over how large the government subsidy to EOSAT should be. The administration has pushed hard to move the financial burden for providing remote-sensing service to the commercial sector. But EOSAT has maintained that the government had made a commitment to providing funds for launches at least through Landsat 6.

Although the plan agreed by NOAA and EOSAT must still be approved by Congress and the White House Office of Management and Budget, both NOAA and EOSAT are confident of success.

Congressman Robert Roe (Democrat,

New Jersey), chairman of the Committee on Science, Space, and Technology, says the new arrangement "removes the uncertainty hanging over our remote-sensing program and enables EOSAT to remain competitive in the international arena". NOAA will provide EOSAT with \$220 million to build the satellite and establish the necessary ground links. But EOSAT has agreed to pay back \$10.8 million from its marketing revenues at the rate of \$2.5 million per year. NOAA also persuaded the Air Force to launch Landsat 6 for \$36.5 million, approximately \$10 million less than the amount originally requested for launch services.

The concern now will be the hiatus in service between the demise of Landsat 4 and 5 and the arrival of Landsat 6. By 1991 the satellites will be in their eighth and ninth years in orbit, and both had a design life expectancy of only three years. EOSAT is husbanding the two satellites' resources, but a gap in coverage now seems inevitable.

Landsat 6 will have a thematic colour mapper similar to those carried on earlier Landsats, with 30-m resolution. It will also have an additional 15-m resolution panchromatic channel easily superimposable on the colour images. This still leaves Landsat's capability behind the 10-m resolution available from SPOT, the French remote-sensing satellite. But if financial support can be found, EOSAT will also put sensors capable of resolving objects to 5 m on board the new satellite.

Another instrument that may get a ride aboard Landsat 6 is the Sea Wide Field Sensor being developed by the National Aeronautics and Space Administration (NASA). Sea WIFS, as the sensor is known, is an improved version of the Coastal Zone Color Sensor that flew aboard Nimbus 7. It measures sea colour changes caused by phytoplankton pigments, chiefly chlorophyll A, and is needed for the global ocean flux study sponsored by the National Science Foundation.

The future of Landsat 7 is less certain. Both Congress and the Reagan administration support an additional satellite. NOAA has awarded contracts to KRS Remote Sensing, a division of Eastman Kodak, the Analytical Sciences Corporation and the Egan Group to conduct comprehensive studies of how a new satellite should be configured. EOSAT is conducting its own study, and will make an independent proposal for Landsat 7. It is clear that the administration will look favourably on proposals that shift even more of the cost of the satellites onto the private sector.

Joseph Palca

ESA's Olympus set to launch a new satellite era

London

FORTY educational institutions from ten countries have been selected to participate in a distance-learning experiment using the pioneering Olympus-1 satellite of the European Space Agency (ESA). British interest in the Olympus programme is riding particularly high: ESA's optional telecommunications programme is one on which the British government is sufficiently keen to spend around £20 million annually. The Olympus programme, which started in 1979, will cost an estimated £400 million, of which Britain will contribute 40 per cent. British Aerospace has taken the lead in the satellite's design and construction. Olympus, which started life under the title L-Sat (for 'large satellite') is aimed at identifying key satellite technologies for the next decade.

Olympus-1, whose launch is scheduled for January next year (some 18 months later than originally planned, although no reason has been given for the delay), is described as a 'technology demonstration', and will be followed by a family of multi-purpose communications satellites that will be among the world's largest and most powerful.

For Olympus-1, a total of 3,000 hours a year will be made available for education, with Britain's share at 1,150 hours for the 23 British institutions selected. The services will be provided free for the first two years, after which a nominal fee will be required. The types of exercise for which Olympus will be used include language learning by establishing multilingual video magazines between nations and live transmission of medical conferences. The largest single interest shown has been in broadcasting films or video. Allocation of time for new business services will be made by ESA. The agency has identified the corporate training and educational services as having the greatest potential for expanding satellite use in the next decade.

Olympus-1, which with a launch mass of 3,300 kg will be the largest-ever multiple communications satellite, will have three payloads. The first will allow transmission of multiple beams in the 12–14 GHz range, for data transmission experiments and multi-location video conferencing; the second will allow experimentation in the 20–30 GHz range, the so-called 'frequency of the future', with possible uses in aircraft communications and electronic news-gathering; and the third will allow direct broadcasting, with two high-powered television channels for broadcasting across most of western Europe. Simon Hadlington

Earnings up

London

BRITISH universities continue to become less dependent on block grants direct from the Department of Education and Science. Provisional figures compiled for the University Grants Committee (UGC), released two weeks ago, show that earnings for research and other services increased by more than £90 million in 1986–87 compared with the previous year to a total of £635 million — a 13 per cent increase in real terms. Of the earnings, 49 per cent came from non-government sources (mainly industry, commerce and charities), compared with 37 per cent in 1981–82. Direct income from the research councils rose by 7 per cent in real terms to £180 million between 1985–86 and 1986–87. During the same period, increases were achieved in income from industry and commerce (up 10 per cent to £68 million), charities (up 23 per cent to £93 million) and short courses for industry and other bodies (up 16 per cent to £47 million). These increases in external revenue mean that in 1986–87 the government's block grants to universities amounted to 55 per cent of their total recurrent income, compared with 77 per cent in 1974–75.

Simon Hadlington

Hope for collaboration — brain meets mind in the midwest

St Louis, Missouri

HAS the time arrived when a synthesis may emerge between basic neurobiology and cognitive science? As the home of the Center for Cellular and Molecular Neurobiology as well as the McDonnell Center for Higher Brain Function, Washington University seems ideally suited to resolve the issue. But even within the two centres there are differing opinions about how soon, if ever, research on the mind and brain can come together under one epistemological roof.

For Gerald Fischbach, director of the molecular centre, the jury is still out on the issue, but he is clearly not optimistic. Fischbach's centre consists of 15 faculty members working on the basic biology of the nervous system: how acetylcholine receptors are anchored to cell membranes, how acetylcholine receptor proliferation is induced at neuromuscular junctions, how cellular proteins are organized to form the cytoskeleton of growing nerve cells. As centre director, Fischbach is naturally enthusiastic about the progress of these efforts. But he has his doubts about how the molecular biologist can contribute to more global issues of brain function.

If there is one area that seems promising for collaboration, it is how molecular-level studies might explain changes in neural wiring. It has recently become possible to study changes in the distribution of synapses on individual cells. Using living dyes, low light level video microscopy and digital image processing, Dale Purves and Jeffrey Lichtman are able to show changes in synapse distribution on peripheral ganglion neurons and neuromuscular junctions over six-month intervals. Extending this technique to neurons in the brain, accompanied by appropriate behaviour tasks, may provide a cellular model for studying behavioural plasticity.

Enthusiasm for collaboration is greater at the Center for Higher Brain Function, particularly in the laboratory of Marcus Raichle. Raichle has been using positron emission tomography (PET) to study cerebral blood flow. By injecting a bolus of water made with a radioactive isotope of oxygen with a 123-second half-life, Raichle has been able to take 40-second 'snapshots' of deep brain blood flow. Collaborations with cognitive scientists Michael Posner and Steve Petersen have been fruitful, providing direct observation of brain areas active during processing of verbal information, confirming that the brain uses a parallel rather than a serial approach to processing lexical information (see *Nature* 331, 585; 1988).

Raichle says that by designing appro-

priate experiments it will be possible to find physiological correlates of complex psychological phenomena such as attention. Localizing brain areas involved in such processes will be the basis for more cellular and even molecular studies of brain function related to behaviour. It is not a question of whether tools of biology will be used by cognitive psychologists, says Raichle, but how those tools can be used.

Neuroscience at Washington University has benefited from the interest taken by the James S. McDonnell Foundation in issues regarding mind and brain. McDonnell, who died in 1980, was the founder of the McDonnell Aircraft Company, now McDonnell-Douglas. But in addition to Washington University's centres, the foundation has decided to start an active search for programmes that will start off the next generation of brain

research. Five panels, supported to the tune of \$1.37 million over two years, will bring together researchers studying perception, memory, attention, emotion and thought to explore whether a research agenda can be established for the field of 'cognitive neuroscience' or if indeed such a field can be said to exist.

Foundation president John Bruer believes the time may be right for pulling together normally disparate fields of neuroscience, developing an approach to asking more unified questions. Bruer says there has been an enthusiastic reception for a summer institute in cognitive neuroscience to be held in Boston this year.

But Bruer says his foundation is prepared to look elsewhere for the synthesis if after two years the panels say the time is not yet right. Fischbach sees the challenge of bringing the two centres at Washington University into greater collaboration as a bellwether for the success of future collaborations. "If it doesn't happen here", says Fischbach, "it isn't going to happen."

Joseph Palca

South African institutes restructured

London

SOUTH Africa's new Council for Scientific and Industrial Research (CSIR) will be officially commissioned on 1 April. The past nine months have seen a programme of rationalization in the largest of the three major research councils in order to cope with a reduction in state funding.

Impetus for the shake-up came after notice from the government that the state subsidy would be fixed at the 1986-87 figure of R234 million (see *Nature* 327, 275; 1987). The CSIR has responded by restructuring its 23 national research institutes into 11 technologically related divisions intended to attract increased contracts from both the private sector and government departments. Fundamental as opposed to applied research will no longer be undertaken by the council's laboratories, but will be the sole prerogative of the universities, funded by the Foundation for Research Development (FRD), the council's semi-autonomous wing, and serviced by the FRD's three national facilities: the South African Astronomical Observatory, the Radio Astronomy Observatory and the National Accelerator Centre.

The CSIR's president, Dr Chris Garbers, said recently that he was "aware that these extensive changes cannot be painless". Eight senior scientists (including two vice-presidents) in their sixties have been retired early, and a further six younger chief directors' posts have been cut. Of the 11 heads of the new research divisions, only one was appointed from outside the CSIR.

The stated aim of the restructuring is to use the country's (pitifully low) research and development funding as efficiently as

possible — this constituted only 0.93 per cent of South Africa's gross national product in 1983-84, the last financial year for which figures are available. The government has made it quite clear that its contribution to research and development will not increase in the foreseeable future, so the CSIR has decided that its salvation lies in orientating itself towards applied research. Of private sector expenditure on research and development, 95 per cent is on 'in-house' projects; only 2.7 per cent goes to universities, and the remaining 2.3 per cent to the research councils (1983-84 figures). In 1986-87, the CSIR earned R87 million (out of a budget of R321 million) from contract research. In the financial year now ending, this is expected to increase to R108 million, most from contracts from state departments. Both the CSIR and the universities seek to increase private-sector funding.

The FRD seems happy to assume responsibility for fundamental research in the hope that its funding will increase accordingly; if that happens, the universities stand to gain as a result of the CSIR's rationalization. In 1986-87 the FRD allocated R42 million to university research, increased to R48 million in the financial year now ending, just below the 16 per cent required to keep abreast of inflation. The FRD requires almost twice its present budget, however, to phase in research funding for all the evaluated researchers in the universities. How much the FRD's budget is to be increased to match its new-found status as custodian of fundamental research remains to be seen.

Michael Cherry

India and the United States agree on vaccine programme

New Delhi

INDIA and the United States have decided to go ahead with the implementation of the Indo-US Vaccine Acting Programme (VAP) despite the controversy stirred up by the signing of this agreement in July 1987 (see *Nature* 328, 287; 1987).

This decision was announced at the end of the first meeting of the Joint Working Group (JWG) co-chaired by Dr A.S. Paintal, director-general of the Indian Council of Medical Research, and Dr Nyle Brady of the US Agency for International Development. The group, consisting of eight scientists from each of the two countries, has been set up to direct the five-year programme against the most important diseases in India: viral hepatitis, diarrhoea, pertussis, animal rabies, respiratory infections and polio.

Originally scheduled for launch in September last year, VAP became bogged down in a controversy because of fears among the Indian science community that it would make India a testing ground for genetically engineered vaccines developed in the United States and that the US government would be the beneficiary of militarily sensitive epidemiological data. Paintal himself was of the view that no vaccines should be tested in India unless they had already been approved for use in the United States, and Dr G.S. Bhargava, a noted molecular biologist, was critical of trying imported recombinant vaccines on people before India has its own production and quality control facilities.

The Indian government's answer was to create an apex committee with Paintal as chairman and Bhargava as one of the members. All VAP activities must be cleared by this committee before submission to JWG for implementation. The government also assured the critics that no American will be directly involved in collection of epidemiological data. Ethical guidelines for conduct of human trials and a framework for sharing the intellectual property arising out of VAP have been spelled out in mutually agreed documents that are currently being scrutinized by the Indian law ministry.

The US team was particularly annoyed at the fuelling of the VAP controversy in India by the row over trials of the Wistar Institute's vaccinia-rabies recombinant vaccine on cows in Argentina. The reasons for the Argentine controversy "were purely political and not scientific", said Professor Frederick Robbins, a JWG member and Nobel laureate from Case Western Reserve University. He hoped India would not abandon trial of this vaccine on stray dogs because of what happened in Argentina. Dismissing specu-

lation that the US interest in VAP was driven by commercial or other motives, he said: "We are in it because it made good foreign policy. But I care a damn whether Indians want to test our vaccines or not".

The JWG has scheduled its next meeting for October in Washington, by which time concrete research proposals will have been received from identified investigators in India and the United States. According to Paintal, the apex committee will examine each proposal from ethical, safety and environmental aspects.

With the hurdles removed, VAP is set for a start, but doubts remain whether the

availability of new vaccines will in itself contribute to the improvement of Indian health. Tuberculosis has been on the rise although BCG vaccination has been part of childhood immunization for 30 years, and tetanus is the principal killer of children in India despite the availability of an effective vaccine. According to Dr Jacob John, a member of JWG and a virologist at the Christian Medical College in Vellore, vaccinated individuals constitute 30 per cent of total paralytic polio cases in the country although the same vaccine has virtually eradicated polio from Europe and the United States. It remains to be seen whether high-technology vaccines would be more effective than protected water supply and improved sanitation in preventing deaths from typhoid and diarrhoea.

K.S. Jayaraman

Change of heart in West German bacterial release debate

Munich

WEST Germany has moved a step closer to approving the limited release of genetically engineered microorganisms. Experts from research and industry spoke out against a proposed five-year moratorium on the release of genetically engineered microorganisms at a parliamentary hearing in Bonn on 2 March. A single biologist backed a moratorium but approved of small releases as long as they could be proved "reversible". The hearing was marked by a lack of polemic, in contrast to the sharp public debate spurred by a release experiment last year in Bayreuth (see *Nature* 328, 568; 1987).

The five-year moratorium was proposed last year in the report of a parliamentary commission convened to assess the risks of genetic engineering. The commission would have allowed plant and animal releases if approved by the Central Committee for Biological Safety (ZKBS), a division of the Federal Health Office. Initially, all the large political parties supported a moratorium.

The Christian Democratic Union (CDU), which shares power with two smaller parties in the ruling Bundestag (parliament) coalition, is now moving towards permitting "smaller experiments" for the purpose of safety assessment, according to CDU Bundestag member Heinrich Seesing. The CDU fears new industries might be damaged by a moratorium but Seesing denied that economics were the party's only reason for its change of heart.

The issue will be dealt with in a sweeping "gene law" due to be completed by the Health Ministry later this year. Under current law, release of genetically manipulated plants, animals or bacteria is allowed only with the permission of the

ZKBS. Technical regulations for release of any genetically engineered organism are expected to be available before the new law is complete, possibly by the end of April. They will specify the criteria by which the ZKBS should judge release applications. No application has so far been approved but at least two new applications are expected this year.

The European Community (EC) is also expected to issue guidelines later this year. The European Parliament is considering a number of proposals, some of which would ban all types of release.

The lone supporter of a moratorium at the hearing was biologist Berndt Heydemann of the University of Kiel. Heydemann said that a five-year ban would not be long enough to determine the impact of the altered organisms on the surrounding ecosystem. Ten to fifteen years of observations would be needed before conclusions could be drawn. Heydemann also pointed out the shortage of funds for risk assessment research in ecology, a concern shared by Wolf-Michael Catenhusen (SPD), chairman of the Bundestag Research and Technology Committee. Catenhusen continues to favour a moratorium.

But the other experts, including ZKBS chairman Werner Goebel, called for a case-by-case review of all release experiments. Köln biologist Jozef St Schell declared that a moratorium would create the false impression that genetically manipulated organisms present an "uncontrollable menace" and that West German institutions are unprepared to oversee and regulate release experiments.

West Germany's Green Party continues to call for a ban on all genetic engineering and a moratorium on release. But the tide seems to have turned. Steven Dickman

CNRS strengthens links with industry

Paris

THE Centre National de la Recherche Scientifique (CNRS), the largest of France's state-funded research bodies, has announced the creation of two major research partnerships with industry. The new ventures, one in immunology and the other in metallurgy, reflect the pattern of co-funding with industrial partners being promoted by the present conservative government.

In the fast-changing field of biotechnology, this kind of partnership has obvious mutual benefits in the effort to maintain international competitiveness. While state-funded researchers gain access to expensive facilities, the industrial partner is able to tap into the latest developments in molecular biology, often arising from basic research. This was the thinking behind the creation of a joint laboratory by CNRS and the pharmaceutical company bioMérieux.

The new laboratory has been given space at the École Normale Supérieure at Lyons, under the direction of Dr B. Mandrand, of the Department of Immunological Research at bioMérieux. The principal objectives of the collaboration are to develop new *in vitro* diagnostic techniques and to simplify and automate immunological reactions. Apart from work specifically aimed at the development of early diagnosis of retroviruses, the group will extend an existing collaboration between the CNRS organic materials laboratory and bioMérieux on the development of new reagents.

The joint laboratory, like others already set up with companies such as Saint Gobain, Elf Aquitaine and Rhône-Poulenc, will be funded equally by the partners, initially for 4 years, but a spokeswoman for CNRS would not disclose the budget set aside for the project.

The new Titanium Research Group, announced by CNRS on 1 March, is more diffuse in its organization. Five university metallurgy departments — three of which have CNRS funding — five industrial companies and the government's military research department, DRET, have joined forces in a FF20 million (£2 million) effort to promote research into titanium and its alloys.

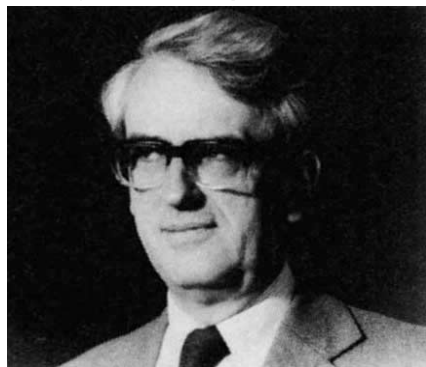
Last year, research minister Jacques Valade announced his backing for new materials research, but the targeting of titanium alloys, such as TA6V, is a strategic decision. The most immediate application of these alloys is in aerospace industries to which France has a long-standing commitment, both as exporter of military aircraft and manufacturer of space hardware within the European Space Agency.

Peter Coles

Phillips defends three-tier model for UK academic science

London

THAT British scientists remain deeply divided about the best prescription to cure the nation's ailing science base was clearly demonstrated two weeks ago at a meeting organized by the Society of Chemical Industry (SCI). The present debate was



Sir David Phillips—sticking to his guns.

ignited last July with the publication by the Advisory Board for the Research Councils (ABRC) of a discussion document entitled *A Strategy for the Science Base* (see *Nature* 328, 280; 1987). The most contentious of the document's 35 recommendations was for differentiation between types of higher education institution engaged in scientific research: type R (with substantial research activity across all fields); type X (with substantial research activity in particular fields; and type T (pursuing the research and scholarship necessary to support teaching but without advanced research facilities). The resulting clamour of opposition from the higher education community has been well documented (see, for example, *Nature* 330, 3; 1987). But at the SCI meeting, ABRC chairman Sir David Phillips was sticking firmly to his guns.

Phillips seems to have a more sympathetic — or perhaps realistic — view than many of his critics of the government's position on the public financing of science. In the House of Commons debate on 29 February, two days before the meeting, Mr Robert Jackson, the higher education minister, reaffirmed the government's view that the problems facing British science will be solved only by more effective management: "It is simply not possible to bridge the gap between the reach of our scientists and their grasp by increasing public expenditure". Similarly, Phillips says the key to the problem is to learn how to manage the system without depending on continuing growth. "Unless we are to lurch from crisis to crisis, we badly need both a deeper understanding of the dynamics and management of scientific research and the resolution and resources to tackle these problems effec-

tively." Furthermore, a year ago Mr Kenneth Baker, the Secretary of State for Education and Science, made clear his view that research should be concentrated in fewer universities or university departments, possibly to the extent that some universities should become primarily teaching institutions (see *Nature* 326, 116; 1987).

While welcoming the fact that the ABRC document had appeared to focus the minds of the academic community more sharply on the long-term problems facing science, Phillips feels that much of the opposition to the 'RTX' proposals is based on misconception.

He accepts that the University Grants Committee is moving some way towards achieving a more streamlined system by its present policy of selectivity, but says that this will not achieve quickly enough the degree of concentration needed. Phillips emphasizes that the RTX categorization was intended to apply to the polytechnics as well as the universities. "Given that not all such institutions are funded for research in the same way at present, there should at least be pause for thought about the potential standing of type T institutions and their role in the developing educational scene." He "regrets" that the RTX proposals were perceived as introducing a hierarchy. "The difference between a type R and a type X institution would have more to do with the range of disciplines in which there was high level research than in the quality and scope of that research. He also stresses that the board did not recommend that differentiation should be imposed from above, "for example by the Secretary of State issuing a list on April 1 1989", but that the institutions themselves should identify their main strengths and future roles.

Phillips decries as a myth the notion that every British university can and does maintain a high level of research activity across a wide range of disciplines. "It is a myth sustained by three factors: first by a belief that university teaching and research are inseparable; second by the opaqueness of the dual support system of university funding; and third by the wish of most universities to retain a presence in as many disciplines as possible."

Responding to Phillips, Professor Ron Hester, of the University of York, contended that the ABRC's proposals offered little hope of attracting the badly needed increase in government support for the science base. "Without that hope the changes envisaged by the ABRC will lead to little advantage for the universities and will not justify the proposed upheaval."

Simon Hadlington

Darwin's "claw", not "paw"!

SIR—On 29 January 1697, Isaac Newton received from Johann Bernoulli two challenge problems designed to humiliate him, and to demonstrate the superiority of Leibnitz's approach to the calculus¹. Newton received the two problems, in particular the famous problem of the Brachystochrone—the curve of quickest descent between two arbitrary points—at 4 P.M. after a hard day at the Tower of London where, as Warden of the Mint, he had for 18 months been supervising the great recoinage. Newton evidently saw the problems as a direct challenge to his prowess as a mathematician and his claim to independent, if not prior, discovery of the calculus; as they were in fact intended, for Bernoulli had strongly implied elsewhere that Newton had borrowed significantly from Leibnitz.

Newton did not go to bed until he had solved the problems, at 4 in the morning. His solutions were despatched, anonymously, the next day. On receiving them, Bernoulli made his famous reputed remark, passed down to generations of British schoolboys as "I recognize the lion by his paw!".

Wishing recently to refresh my memory on this for a college core course lecture, I discovered to my astonishment that Bernoulli's Latin has been badly translated. When corrected, a completely different image emerges than that conventionally implied.

Westfall², describing Bernoulli's response, reports his remark differently: "Disabused on Newton's skill in mathematics, Bernoulli recognized the author in the authority the paper displayed — 'as the lion is recognized from his print', in his classic phrase.¹⁰⁸" Footnote 108 reads: "In Latin, of course: 'tanquam ex ungue leonem'. It is indicative of Newton's pride in his solution that he believed l'Hôpital did not succeed without help."

While paws can indeed leave prints, they are obviously not the same word, so the question naturally arises whether Westfall's translation, "print", is to be preferred to the time-honoured "paw". A hint that Westfall's translation is suspect is already contained in the fact that he has implicitly changed the voice in translation as "leonem" is in the accusative.

The English language has served to confuse the meaning to be attributed to "ungue", in words ultimately derived from that root. For *Webster's Dictionary* tells us that, while the unguiculata are mammals possessing nails or claws, the ungulata are hoofed mammals as distinct from the preceding; yet, in heraldry, unguled means having hoofs or claws of a heraldic tincture different from that of the body.

It is however clear that lions do not have

hoofs; and the straightforward meaning of the word unguis (ungue being the ablative) is "claw or nail", as confirmed by many Latin dictionaries. My colleague Professor M.K. Gamel informs me that the text would require pede (from "pes") had "paw" been intended, while Westfall's "print" would require "vestigium" (from "vestigium").

Thus we see that Bernoulli did not 'recognize the lion by his paw' but rather by his *claw*. The problem, designed to taunt the old British lion in his den at the Mint, where he had been thought to be out of action for some time, had been raked over by Newton's claws and tossed back, imperiously and disdainfully. This seems much more in keeping with the kind of surprised reaction Bernoulli must have had (considering the intent behind the problem), than the almost unbelievably noble sentiments implied by the long-transmitted translation.

That would be the end of the matter but for a rather telling use of the word unguem in Lewis and Short's *Latin Dictionary*². Unguem appears in many proverbial phrases, several involving references to fine work. However, under this heading they also include "Cum medium ostenderet unguem, i.e. *showed utmost derision, the greatest contempt* (because the middle finger was regarded as indecent)". Could it be that, given the way the problem was almost contemptuously tossed back to him, Bernoulli felt, as expressed in the modern American idiom, "My God, the old lion has given me the finger!" (tanquam)? This interpretation suggests that, just as much of Shakespeare's earthy and bawdy language was once watered down for later consumption by youthful audiences, so some of the visceral feelings with which these great mathematicians sometimes confronted one another have been sanitized for posterity.

JOHN FAULKNER

Lick Observatory,
University of California,
Santa Cruz, California 95064, USA

1. Westfall, R.S. *Never at Rest* 581–583 (Cambridge University Press, 1980).
2. Lewis, C.T. & Short, C. *A Latin Dictionary* 1931 (Clarendon, Oxford, 1962).

An ethical dilemma

SIR—You recently published a letter¹ in which I argued that progress in AIDS research is being delayed because the present system of funding is inhibiting collaboration between researchers. I reported that after I had responded positively to requests for samples of my cDNA recombinants from various US laboratories engaged in AIDS-related research, my research grant applications were rejected

by three major funding organizations (US and Canadian). Specifically, a Canadian organization gave as a major reason for refusing funding the fact that I had shared my recombinants.

I now have the official summary statement from the National Institute of Health (NIH), giving reasons why that organization rejected my application. One of the criteria used to assess applications from foreign laboratories is whether there is anything unique about the laboratory that could not be duplicated in the United States. The summary answers this by stating: "The special opportunities that exist in Dr Forsdyke's laboratory may be those of his identification and possession of unique cDNA clones; however, these clones are apparently now available to the investigators in the United States".

The lesson from this is very clear. If foreign investigators intend to submit grant applications to the NIH, they should not respond to requests from US laboratories for unique research materials.

But does the left hand know what the right is doing? NIH have recently reissued their 1984 policy statement on the "Distribution of Newly Developed Materials", which says: "Restricted availability of these materials can impede the advancement of basic research and the delivery of medical care to the nation's sick".

The present system of research funding appears to be based on the premise that aggressive competition between researchers advances medical progress optimally. This is manifestly wrong². Suggestions for reform that would decrease competition and enhance collaboration between researchers^{3,5} have been ignored. If research administrators will not put their own house in order, then such order should be legislated.

D.R. FORSDYKE

Department of Biochemistry,
Queen's University,
Kingston, Canada K7L 3N6

1. Forsdyke, D.R. *Nature* **328**, 662 (1987).
2. Ubell, R. *Nature* **294**, 28 (1981).
3. Osmond, D.H. *J. Neurobiol.* **14**, 95–112 (1983).
4. Forsdyke, D.R. *Med. Hypothesis* **11**, 141–145; 147–156 (1983).
5. Forsdyke, D.R. *Nature* **312**, 587 (1984).

Wrong bird

SIR—The wagtail depicted feeding a young cuckoo (*Nature* **331**, 19; 1988) may be considered unfortunate enough to be lumbered with this unwholesome 'offspring'. To further mislabel this grey wagtail (*Motacilla cinerea*) as a yellow could cause a serious identity crisis. The cuckoo, admittedly, does not look too bothered about the confusion.

ANDREW O. M. WILKIE

MRC Molecular Haematology Unit,
Room 7501,
John Radcliffe Hospital,
Oxford OX3 9DU, UK

Global ozone trends reassessed

A new study reported in this issue says that ozone is being depleted globally. But the extent of the depletion is debatable — and has been debated.

IN a paper on page 219 of this issue, Donald Heath now of the National Oceanic and Atmospheric Administration (NOAA), reports large decreases in global ozone concentrations measured using the Solar Backscatter Ultraviolet (SBUV) instrument on the atmospheric research satellite Nimbus 7; the decreases are much larger than expected from theoretical calculations. A working group set up by the National Aeronautics and Space Administration (NASA) has exhaustively studied Heath's data, other ozone measurements and photochemical models of the atmosphere, and will have reported its findings at a press conference in Washington DC this week. The group concludes that there is a serious drift in the calibration of SBUV, but that even after accounting for the drift, the global depletion of ozone is real. In the light of the NASA study, it is clear that Heath's uncorrected trends of global ozone depletion are too large. When the data are corrected for calibration drift, the trends agree reasonably well with current views of how anthropogenic emissions perturb the chemistry of the atmosphere.

Heath's data suggest that there has been a large decrease in global ozone since about 1978 and that an Arctic 'ozone hole' has developed, similar to, but less pronounced than the Antarctic 'ozone hole'. The decreases reported by Heath are considerably larger than those predicted by photochemical models of the effect of chlorofluorocarbons (CFCs) and solar-cycle variations on the atmosphere. K.P. Bowman (*Science* **239**, 48-50; 1987) finds trends of similar magnitude using archived data from the Total Ozone Mapping Spectrometer (TOMS) on Nimbus 7.

Heath's report to a Congressional Committee in 1986 of a preliminary version of his findings caused NASA, in collaboration with other agencies, to set up the Ozone Trends Panel for a critical evaluation of all available ozone data.

The discovery of the Antarctic ozone hole was a surprise (see *Nature* **315**, 207; 1985). It had not been predicted by any photochemical model, and there was no immediately obvious explanation for it. The primary importance of the discovery was the realization that if the chemistry of such gross changes is not understood, other subtle yet significant changes going on elsewhere might also have been missed. The SBUV/TOMS trend could have been one of these significant

changes, which is why the Ozone Trends Panel (OTP) was set up.

The available reasonably long-term data set examined by the panel includes the surface network of Dobson spectrophotometers, the SBUV and TOMS instruments and the Stratospheric Aerosol and Gas Experiments, SAGE I and SAGE II, on the Applications Explorer Mission-B (1979) and the Earth Radiation Budget Satellite (1984), respectively.

At the outset, the panel found all the ozone data to be unsatisfactory. The archival Dobson data suffer from quality-control problems, such as unrecorded recalibrations of individual instruments. The SBUV and TOMS data suffer from the degradation of their common calibration mechanism, a diffuser plate reflecting solar radiation into the instruments to provide a full-scale measurement. When the study began, the SAGE II measurements were still being validated, while the SAGE I data had been processed by a different algorithm (and was reprocessed for the panel).

Heath points out a drift of about -0.4 per cent per year of the SBUV total column ozone with respect to the Dobson network. The panel found that by making different, but reasonable, assumptions about the degradation of the diffuser plate, this drift can be eliminated. They conclude that the ozone trend in the archived SBUV data set, as reported by Heath, is overestimated, and that the uncertainties are larger than that stated by the data-processing team.

A significant fraction of the Dobson data was re-analysed for the panel, taking into account instrument calibration changes and station reliability. These data do not give global coverage, but are believed to be reliable, and have been used to renormalize the TOMS and SBUV data by guiding the selection of a diffuser degradation model. The renormalized TOMS data show a global decrease of 2.6 per cent from November 1978 to October 1985. After 1985, there is a hint of an increase in the Northern Hemisphere, but depletion continues in the Southern Hemisphere with negligible global change.

Model calculations to include the effects of CFCs and solar-cycle variations were carried out for the panel by four different research groups. They show that the solar cycle should have had a comparable or greater effect on total ozone than that of the trace gases over the period

1979-85. Calculated changes are broadly consistent with the observations in the Northern Hemisphere, but not in the Southern Hemisphere, perhaps because the (unmodelled) Antarctic ozone hole is offsetting the solar-cycle effect throughout the Southern Hemisphere.

For the period up to the next solar maximum, the OTP predicts that the increase in ozone resulting from increased solar ultraviolet radiation will mask or even reverse the decrease caused by increasing concentrations of CFCs. This implies that, after the solar maximum, the decrease will continue with renewed vigour.

Data from SAGE I and SAGE II during the period 1979-85 show a maximum depletion of about 3 per cent in ozone density at an altitude of 40 km, with no change at 50 km. A re-analysis of the SBUV vertical profile data, using the diffuser plate model that eliminates the bias between SBUV and Dobson data, brings the profile trend into better agreement with the model calculations. For the same period as SAGE, the maximum depletion is around 7 per cent at 40 km, decreasing to near zero at 50 km.

The model calculations all show maximum global depletion around 40 km, varying from 5 to 12 per cent for different models. Of this, 4-9 per cent is response to CFCs, and 1-3 per cent is the response to reduced solar ultraviolet output approaching the solar minimum.

Important indirect support for the photochemical models can be found in stratospheric temperature trends. It seems that the photochemical models and the latest measurements of atmospheric ozone trends are beginning to converge, and are giving broad agreement globally, although the Antarctic ozone hole is still not well-enough understood to be included in models. The error bars for both measurements and models are still large, but quantitative evidence for the effects of CFCs on ozone in the atmosphere is now much more convincing. The processing and release for study of the data from the SBUV-2, launched in 1984 by NOAA on an operational meteorological satellite, are eagerly awaited. These data should settle any remaining doubts about what is really happening to the ozone layer on the global scale.

Clive Rodgers

Clive Rodgers is a member of the Ozone Trends Panel and is at the Department of Atmospheric Physics, Oxford University, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK.

Vertebrate palaeontology

Embryos in dinosaur nests

David B. Norman

EVENTS such as the discovery of dinosaur nests containing preserved embryos are exceedingly rare, so the report of such a discovery by Horner and Weishampel on page 256 of this issue¹ is of considerable interest. The new discovery is from the sediments of the Two Medicine Formation (75 million years old) around Egg Mountain, Teton County, Montana. Horner^{2,3} has previously reported the discovery of eggs, the structure and arrangement of nests and even colonial nesting sites of mainly hadrosaurian (large, 8-m-long, bipedal, herbivorous 'duck-billed') dinosaurs belonging to the genus *Maiasaura*. These discoveries provide hitherto unimagined insight into the social life of these types of dinosaur.

Horner's work reveals, for example, that the nests of these dinosaurs were large, vegetation-lined mounds, spaced at fairly regular intervals in dense colonies, and that such sites were returned to and refurbished seasonally. It is also possible that hatchling *maiasaurs* tended to stay in or around the nest for some time, and were presumably brought food by their parents. The similarity between this type of behaviour and that of modern birds (and even some crocodiles) is striking.

Horner and Weishampel in this issue¹ report the first discoveries of remains of *Maiasaura* embryos and of the hypsilophodontid *Orodromeus*. The latter (a small, 2.5-m-long, bipedal herbivore) has already been referred to vaguely, but until now had not been named or described. Horner and Weishampel found the *Orodromeus* remains in a clutch of 19 unhatched eggs (see figure) and the *maiasaur* remains as scattered embryonic bones within a nest mound.

Careful analysis of thin sections of the skeletal remains of both these dinosaurs reveals some curious differences. The joint (epiphyseal) surfaces of the limbs in the smaller *Orodromeus* are reported to be well formed and composed of smooth calcified cartilage; by contrast, those of the hadrosaur *Maiasaura* are poorly finished off. The authors suggest that this difference is because the smaller (2.5-m adult length) *Orodromeus* was a precocious developer, which could scamp about in search of food almost immediately after hatching. *Maiasaura*, whose embryos were destined to become adults of 8 m or more in length, was (as Horner had surmised previously³ using a different argument) a late developer, spending longer in the nest and being fed by its parents.

Additional points that emerge from the

new study¹ include the observations that certain bony processes for muscle attachment do not seem to ossify until the muscles become active after the dinosaur has hatched; and that the denticles (coarse projections), which run down the edges of the teeth in more mature *Orodromeus*, are completely lacking during the early ontogenetic stages.

Reports of juvenile or embryo remains of fossil species are very rare. After all,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Reconstruction of *Orodromeus makelai*, the hypsilophodontid dinosaur embryo found by Horner and Weishampel¹ in the Upper Cretaceous sediments of western Montana. The egg is 17 cm long. (Reconstruction by Matt Smith; courtesy of the Museum of the Rockies.)

just about every factor is against the discovery of such forms. Juvenile organisms are inevitably small and fragile, so their remains tend to be destroyed if they die prematurely, either by the natural processes of decay or by scavengers. As a result, there is an inevitable bias against remains of juveniles being preserved in the fossil record. This is compounded by the fact that not only are they difficult to find in the field, but even if such small and often fragmentary remains are recovered, it is difficult to distinguish between genuine juvenile specimens and the

remains of mature individuals of naturally small species.

Despite these problems, several such finds have been recently reported. Currie⁴ has briefly alluded to the discovery of similar hadrosaur nest sites in southern Alberta. This site is of about the same age as Horner and Weishampel's site, and was reported to include the remains of hadrosaur eggs with embryos. No detailed description has yet been produced, but it will be interesting to see how Currie's finds compare with those from Montana described in this issue.

Slightly further afield both stratigraphically and geographically, Sander has very recently reported⁵ a nothosaur embryo from Middle Triassic (230 million-year-old) sediments in the southern Alps. *Neusticosaurus*, a small (0.2–0.4-m-long) lizard-like creature, is known from hundreds of skeletons preserved in the bituminous shales of Monte San Giorgio. Sander's survey of this material reveals the existence of a probable embryo among the collections in Zurich. The specimen suspected of being an embryo is only 51 mm long and, unlike all other skeletons in the collection, is curled up with the head close to the tail (a typical position for embryos). In addition, the skeleton is poorly ossified and the skull is relatively large when compared with the proportions of the rest of the body, and the eye is very large. Comparisons with living reptiles indicate that the embryo was only about two-thirds of the expected hatchling size at the time of death.

Compared with the dinosaur embryos discovered by Horner^{1–3} and by Currie⁴, the nothosaur embryo described by Sander⁵ does not provide the same degree of insight into the behavioural repertoire of the species. Indeed, there still remain several fundamental problems of interpretation. It is not firmly established whether *Neusticosaurus* bore live young or laid eggs. Ovipary is supported by the facts that no gravid females have been found among the collections, and that these forms could move around on land; they were not obliged to be viviparous as were some of their more highly aquatic cousins such as the ichthyosaurs⁶. On the other hand, the specimen could be an aborted embryo, because the sediments in which these fossils are found are unquestionably marine (with the coast several kilometres away). But how could a shelled egg have been transported so far out to sea?

As Horner has shown, terrestrial egg-

1. Horner, J.R. & Weishampel, D.B. *Nature* **332**, 256–257 (1988).
2. Horner, J.R. & Makela, R. *Nature* **282**, 296–298 (1979).
3. Horner, J.R. *Nature* **297**, 675–676 (1982).
4. Currie, P.J. *J. vert. Paleont.* (Abst.) **7**, 15A, No. 30.
5. Sander, P.M. *Science* **329**, 780–783 (1988).
6. Romer, A.S. *Vertebrate Paleontology* 3rd Edn (University of Chicago Press, 1966).
7. Andrews, R.C. *Natural History of Central Asia* **1** (1932).
8. Brown, B. & Schlaikjer, E.M. *Ann. N.Y. Acad. Sci.* **XL** (3), 133–265 (1940).

layers such as dinosaurs can leave important clues about their reproductive behaviour, and there is undoubtedly much more to be learned from these recent finds. But the new data do have implications for previously known material. Two immediate examples spring to mind: abundant and well-preserved nests of the dinosaur *Protoceratops* were discovered^{7,8} in the Gobi Desert by the American Asiatic

expeditions between 1922 and 1925; and abundant dinosaur eggs attributed to the sauropod *Hypselosaurus* are known from Aix-en-Provence in France. Both these well-known localities could benefit substantially from more detailed studies. □

David B. Norman is at the Nature Conservancy Council, Peterborough PE1 1UA, UK and is a research associate at the University Museum, Oxford, UK.

Cell biology

Oncogenes and cell control

Matilda Katan and Peter J. Parker

NEW insights into cellular control mechanisms and oncogene function are provided this week by two papers on pages 269 and 272 of this issue^{1,2} describing the predicted amino-acid sequences of two apparently unrelated proteins. One report, by Mayer *et al.*², examines the nature of the gene in an avian sarcoma virus that confers the ability of this defective retrovirus to transform cells. The companion paper, by Stahl *et al.*¹, describes the cloning of a complementary DNA encoding a phosphatidylinositol-specific phospholipase C. Surprisingly, the predicted sequences of these two polypeptides share regions of homology that are also present in tyrosine kinases, suggesting some conserved functional domains in these apparently diverse proteins.

Mayer *et al.*² report the amino-acid sequence of the avian sarcoma virus CT10. The single long open reading frame encoded by this defective retrovirus produces a gag-fusion polypeptide p47^{gag-crK} of relative molecular mass 47,000, of which

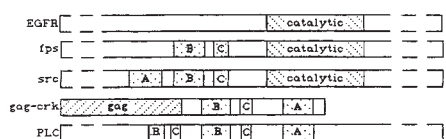


Fig. 1 Location of the A, B and C regions (see text) in: EGF receptor (EGFR); cytoplasmic tyrosine kinases (c-src and c-fps); the CT10-transforming protein p47^{gag-crK} (gag-crK); and PLC-148 (PLC).

232 amino acids are non viral (non-gag) in origin. As observed with several other transforming retroviruses, CT10 induces an increase in phosphotyrosine content in infected target cells. In contrast to many of these previously described retroviruses, however, the transforming protein encoded by CT10 is similar to the non-catalytic domain of the cytoplasmic but not receptor members of the family (see ref. 3 for review). Three distinct stretches of amino acids in p47^{gag-crK} seem to be similar to regions of the prototype cytoplasmic tyrosine kinase c-src (see Fig. 1). Similar regions are conserved in other

cytoplasmic tyrosine kinases, though the 'A-domain' does not appear in all of these proteins (for example, c-fps; Fig. 1). Indeed, the transposition of A, B, C in c-src for B, C, A in p47^{gag-crK} suggests that these domains could have independent functions (B+C and A). Previous work on tyrosine kinases has shown that one of the conserved non-catalytic domains, termed SH2, has a regulatory role perhaps in directing specific interaction with some cellular component(s); this SH2 domain encompasses B+C sequences. The intriguing question posed by the data of Mayer *et al.*² is how the expression of a regulatory-like domain confers transformation capability. Surprisingly, a clue to the answer comes from the companion paper¹ by Stahl *et al.*, on the sequence of a bovine phosphatidylinositol-specific phospholipase C (PLC-148).

There has been much interest in phospholipase C enzymes specific for inositol-containing phospholipids (PI-PLC) since it was realized that these enzymes mediated production of the second messengers inositol 1,4,5-trisphosphate and diacylglycerol (see ref. 5 for review). From studies of PI-PLC it has become apparent that there are several species of this enzyme, differing in molecular properties and immunologically unrelated^{6,7}. Stahl *et al.*¹ describe the first sequence of a mammalian phosphatidylinositol-specific phospholipase C. The predicted amino-acid sequence for this polypeptide derived from complementary DNA cloning reveals that, like p47^{gag-crK}, it possesses regions homologous to the A, B and C regions of c-src. Interestingly, like p47^{gag-crK}, in PLC-148 these regions are transposed, and in the case of B and C duplicated (see Fig. 1).

It is not clear how these conserved domains fit into the overall structure of PLC-148 since there is as yet no definition of the regulatory and catalytic domains of this enzyme. Nevertheless, the homology itself tends to suggest that this central region of the PLC performs a regulatory function. The implication is that several tyrosine kinases and PLC-148 share some common regulatory mechanism and that

expression of p47^{gag-crK} is interfering with or mimicking one, or perhaps both, processes.

The regulation of PI-PLCs has yet to be rigorously defined, and the multiplicity of such activities could infer that, like the tyrosine kinases, there are different means of regulation. It is clear that PI-PLC activity is increased in cells following addition of various agonists, and there is circumstantial evidence that G-proteins are involved in receptor coupling (see ref. 8). Thus, it is tempting to speculate that G-proteins, or perhaps a new class of regulatory molecules, control both PI-PLC and tyrosine kinases (see Fig. 2). Although the simplicity of such a model is appealing, it would not account for the consequences of p47^{gag-crK} expression. As Mayer *et al.* discuss, one interpretation of the effects of p47^{gag-crK} would be that it functions to deplete negative regulatory factors, leading to increased tyrosine kinase activity. Thus a minimal

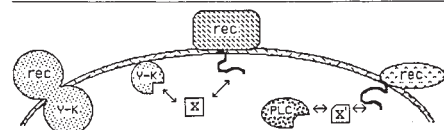


Fig. 2 Model for the regulation of cytoplasmic tyrosine kinase (Y-K) and phosphatidylinositol-specific phospholipase C (PLC). The common regulatory domain in the Y-K and phospholipase C interact with the same class of regulatory proteins (X, X') linked to receptors (rec). This regulation is distinct from the Y-Ks that are part of the cytoplasmic tail of the receptors and that do not possess this conserved domain (left).

requirement would be to include both positive and negative interactions with these conserved regulatory domains to account for the transforming ability of CT10 (perhaps A and B+C define such separate recognition sites for positive and negative effectors). This would not be such a flight of fancy, as the prototype for G-protein action, receptor-adenylate cyclase coupling, involves both positive (G_s) and negative (G_i) protein effectors (see ref. 9). Whatever the intricacies of these systems, the homology between the tyrosine kinases, p47^{gag-crK} and PI-PLC highlights the need to elucidate the factors or polypeptides that interact with these regulatory proteins. □

1. Stahl, M.L., Ferez, C.R., Kelleher, K.L., Kriz, R.W. & Knopf, J.L. *Nature* **332**, 269–272 (1988).
2. Mayer, B.J., Hamaguchi, M. & Hanafusa, H. *Nature* **332**, 272–275 (1988).
3. Hunter, T. & Cooper J.A. *A. Rev. Biochem.* **54**, 897–930 (1985).
4. Sadowski, I. *et al. Molec. cell. Biol.* **6**, 4396–4408 (1986).
5. Berridge, M.J. & Irvine, R.F. *Nature* **312**, 315–321 (1984).
6. Ryu, S.H. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6649–6653 (1987).
7. Katan, M. & Parker, P.J. *Eur. J. Biochem.* **168**, 413–418 (1987).
8. Litosch, I. & Fain, J.N. *Life Sci.* **39**, 187–194 (1986).
9. Gilman, A.G. *A. Rev. Biochem.* **168**, 413–418 (1987).

Matilda Katan and Peter J. Parker are at the Ludwig Institute for Cancer Research, Courtauld Building, 91 Riding House Street, London W1P 9BT, UK.

Cuprate superconductors

Old behaviour in new materials

Timothy R. Dinger and Stanley W. Tozer

SINGLE crystals of the copper-oxide (cuprate) superconductors, including those of the lanthanum- and yttrium-barium-based materials reported last year and the newer bismuth- and thallium-based materials (see the recent discussion by Forgan and Greaves in News and Views¹), all have remarkably similar behaviour. Solid-state theorists are capitalizing on this similarity as they try to identify an alternative to the traditional 'BCS' theory to explain the novel superconductivity^{2,3}. These similarities in charge-transport, magnetic and optical properties undoubtedly arise from the close structural resemblance between the cuprates at the molecular level. Takagi *et al.* report, on page 236 of this issue⁴, measurements of the transport, magnetic and optical properties of single crystals of $\text{Bi}_2(\text{Sr,Ca})_3\text{Cu}_2\text{O}_x$ that match those of $(\text{La}_{1-x}\text{Sr}_x)_2\text{CuO}_4$ and $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ very closely.

Single crystals of the cuprate superconductors all show a strong anisotropy in the response of their transport properties to applied magnetic fields. The most obvious anisotropy is in the temperature dependence of the upper critical field (the applied magnetic field that quenches superconductivity in the crystal). A magnetic field has a much more pronounced and deleterious effect on charge transport if applied perpendicularly to the Cu-O conduction layers than if applied in the plane.

The two-dimensional structure is also evident in the temperature dependence of the resistivity above the transition temperature. Although Takagi *et al.* report⁴ only the in-plane resistivity of $\text{Bi}_2(\text{Sr,Ca})_3\text{Cu}_2\text{O}_x$, this value agrees in both magnitude and slope (decreasing linearly with decreasing temperature) with that of $(\text{La}_{1-x}\text{Sr}_x)_2\text{CuO}_4$ and $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$.

The *c*-axis (out-of-plane) resistivity of the latter compounds, in contrast, is substantially higher at room temperature and varies inversely with temperature. A comparison of the *c*-axis resistivity of $\text{Bi}_2(\text{Sr,Ca})_3\text{Cu}_2\text{O}_x$ with that of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ would be interesting, as the Cu-O planes are known to be more isolated in the new superconductor.

This anisotropy of superconductivity in the cuprates presents formidable problems. Materials scientists are faced with the task of favourably aligning these layered materials to optimize the effect of their directionally varying properties. For instance, a superconducting solenoidal magnet will require highly textured cuprates to ensure that the maximum field produced along the axis of the magnet does not destroy the superconductivity in unfavourably oriented crystals. Such

textured materials have been produced as thin films of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$, which have superior current-carrying capacities⁵, but the use of these materials in practical designs will be difficult.

The similarities identified between these cuprates should help to elucidate the fundamental mechanism responsible for their superconductivity. It is also hoped that they will lead to materials with even higher, more practical superconducting transition temperatures. To this end, Takagi *et al.* note⁴ that the increasing superconducting transition temperature of these three materials correlates well with increasing charge-carrier concentration. It is expected that other fundamental trends in behaviour will develop as better data are obtained. □

1. Forgan, E. M. & Greaves, C. *Nature* **332**, 14–15 (1988).
2. Anderson, P. W. & Zou, Z. *Phys. Rev. Lett.* **60**, 132–135 (1988).
3. Kivelson, S. *Phys. Rev. B* **36**, 7237–7240 (1987).
4. Takagi, H. *et al.* *Nature* **332**, 236–238 (1988).
5. Chaudhari, P., Koch, R. H., Laibowitz, R. B., McGuire, T. R. & Gambino, R. J. *Phys. Rev. Lett.* **58**, 2687–2690 (1987).

Timothy R. Dinger and Stanley W. Tozer are at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA.

Atmospheric chemistry

Increased tropospheric ozone

S.A. Penkett

In a careful re-examination of a data set on ambient ozone concentrations measured around the turn of the century, Volz and Kley conclude, on page 240 of this issue, that ozone levels in the lower atmosphere (troposphere) in Europe may have doubled over the past 100 years. This finding is as remarkable as the observation of a hole in the stratospheric ozone layer over Antarctica (see page 202) and potentially is just as consequential.

Global weakening of the stratospheric ozone shield is obviously detrimental to life on Earth, but we should not congratulate ourselves that it may be partially

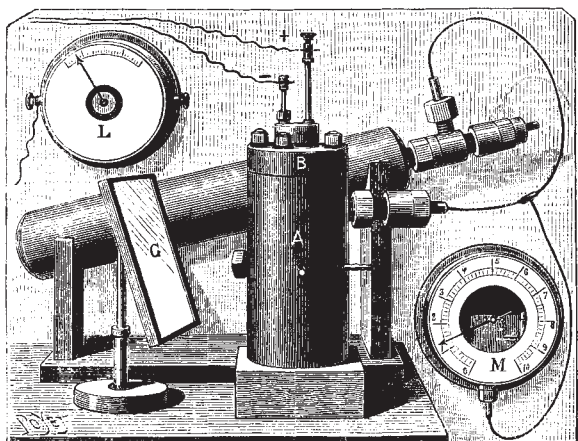
compensated in the Northern Hemisphere by the presence of additional ozone in the troposphere. Such a redistribution is totally undesirable, because ozone in the troposphere is an efficient greenhouse gas, second only to carbon dioxide in importance, and it is highly toxic to plants at concentrations only slightly elevated above ambient levels. Perhaps of more significance in the long term, photolysis of ozone yields hydroxyl radicals which chemically control the trace-gas composition of the atmosphere. Unfortunately, high concentrations of ozone are not the desirable seaside tonic they were once supposed to be.

Interest in atmospheric ozone is not new. It was shown to be a common trace gas in the last century by Andrews (Queen's University, Belfast). Many professional people were then aware of its presence, believing it could act as a disinfectant against disease, and conducted measurements with test papers developed by the German chemist, Schönbein, to determine its behaviour. The database is extensive but, unfortunately, unreliable. In 1876, however, Albert-Levy started a series of measurements at the Montsouris Observatory on the south-west outskirts of Paris using a much more quantitative method, namely iodine-catalysed oxidation of arsenite in neutral aqueous solution. Ozone first oxidizes I^- to I_2 , which in turn

100 years ago

A difficulty often experienced in laboratories is how to raise a body to a high temperature while surrounded by a gaseous atmosphere under considerable pressure. The apparatus makes it possible to bring bodies to a temperature approaching that of the fusion of platinum. A, mass of steel with cylindrical bore, with its stopcock B; G, mirror permitting the reaction to be seen; M, manometer; L, amperemeter.

From *Nature* **37**, 471; 15 March 1888.



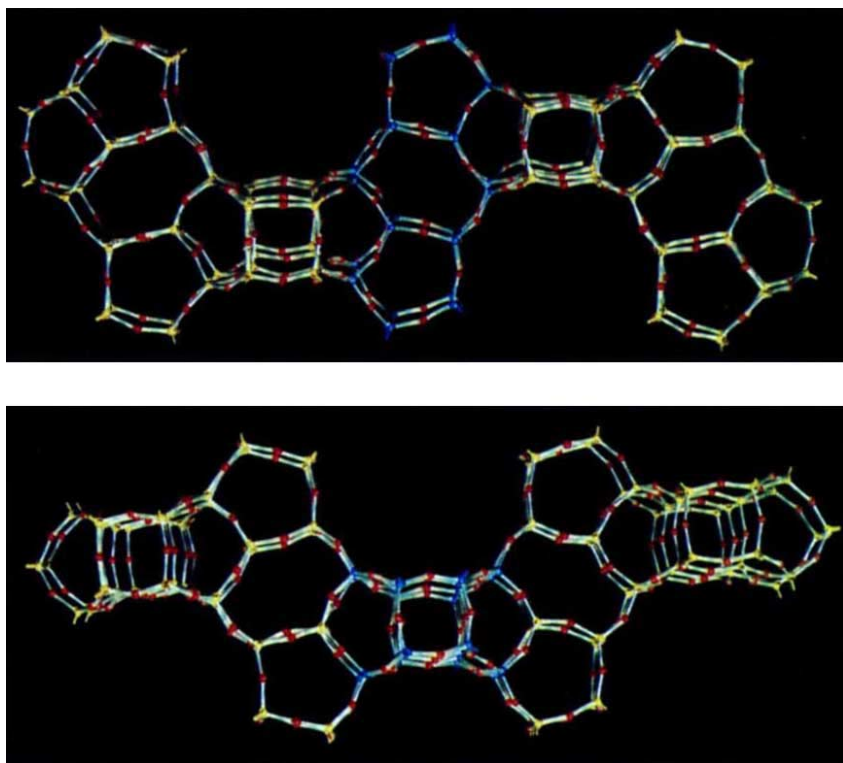
reacts with AsO_3^{3-} , oxidizing it to AsO_4^{3-} , the I_2 being converted back to I^- . The amount of ozone collected in the course of a day in a bubbler system containing these reagents can be ascertained from back titration of the remaining AsO_3^{3-} with I_2 .

The method is precise and agrees well with the modern ultraviolet-absorption method of measuring ozone. There are difficulties with other gases in the air, notably sulphur dioxide which depresses the real ozone signal. The data collected are sufficiently detailed, however, that allowances can be made for this. The results, according to Volz and Kley in this issue, suggest that in clean air from a southwesterly direction, the yearly-average ozone concentration was approximately 10 parts per billion by volume (p.p.b.v.) in the period 1876–86. A recent data set (Feister, U. & Warmbt, W. *J. atmos. Chem.* 5, 1–21; 1987), however, shows that at Cape Arcona on the Baltic coast, the yearly average concentration in 1983 was more than 20 p.p.b.v.. These latter results, of course, were obtained by different workers at a different location. This is not likely to be critical, though, because other similar observations of ozone have been made elsewhere (see Derwent, R. G. (ed.) *Ozone in the United Kingdom*; Dept of the Environment, 1987).

A further interesting point is that the seasonal variation observed in the last century is completely different from that observed now. A small spring maximum observed between February and May has been replaced by a much stronger maximum which peaks in May and extends well into the summer. Ozone concentrations are now much higher all year than were observed 100 years ago. Whereas the large summer values could be associated with regional photochemical air pollution, the yearly increase can be interpreted only in terms of a very substantial perturbation of the ozone budget throughout the surrounding troposphere.

The possibility that an extensive increase in tropospheric ozone results from ever larger emissions of the precursor gases, hydrocarbons and oxides of nitrogen (NO_x), is now obvious to many atmospheric chemists, but confirmatory data have been hard to find. The worrying aspect of the work of Volz and Kley is that the perturbation is already large, and is likely to become larger. Crutzen, for instance, calculates (*Scope* 21, 67–112; 1983) that the lower atmosphere can generate almost an order of magnitude more ozone than it does now, the amount being directly proportional to the concentration of NO_x in clean air.

One of the main sources of NO_x and of many of the reactive hydrocarbons which escape into the atmosphere and subsequently form ozone, is the motor car. Emissions from this source raise the levels of ozone in the summer in the planetary



ZEOLITES are ubiquitous in the petrochemical industry, being used as highly efficient catalysts, sorbants and molecular sieves. The advantageous properties of these aluminosilicates derive from their open-network ('honeycomb') structure. This allows molecules to diffuse easily over the extensive internal surfaces to widely dispersed active sites. On page 249 of this issue, M. Treacy and J. Newsam describe how they painstakingly unravelled the structure of one of the most complicated zeolites — zeolite- β , which was first patented by Mobil in 1967. The structure consists equally of two polymorphs, termed A and B, and it is the extensive intergrowth of these that frustrated crystallographers for so long. The structures above are of polymorph A seen from two perpendicular directions (B not shown). The network of silicon (yellow and blue circles) and oxygen (red circles) atoms defines the impermeable walls of the zeolite. The blue areas highlight the building unit. The conducting channels are formed by the larger rings that are seen only in part in these models, above and below the walls. Treacy and Newsam find that these pores are built from 12-membered rings and, more importantly, that they build up a three-dimensional network which accounts for the high porosity of this zeolite. Although polymorph B has a centre of inversion symmetry (see Fig. 2 on page 250), polymorph A above does not. The resulting 'corkscrew' channels (which spiral around the walls shown), and their asymmetric intersections, could be used to catalyse reactions favouring molecules of a particular 'handedness'. (Photograph courtesy of C.R. Forrester and E.A. Brozyna.) □

boundary layer (up to about 1,500-m altitude) in the form of photochemical smog. Emissions continue in the winter, however, and not all are consumed in the formation of local pollution, allowing some to be dispersed into the background atmosphere. Much of the extra ozone observed in present-day measurements will have been formed when this dispersed pollutant reacts photochemically over a much wider area than the original source. The impact of NO_x emissions is clearly seen on a wide scale in the large increase in nitrate levels observed in Greenland ice over the past two or three decades. Interestingly, the limiting substances for more extensive ozone formation in more localized pollution episodes are the hydrocarbons. Emissions of both these types of substances should therefore be

restricted.

The adoption of measures to remove the problem of tropospheric ozone will be very difficult, probably more so than in the case of stratospheric ozone. The expanding world economy will generate more emissions, particularly of nitrogen oxides which are a mark of an 'advanced' economy. This will probably cause the source areas to spread and intensify ozone production, at least in the Northern Hemisphere. Attempting to place controls on emissions, particularly of nitrogen oxides, will be difficult technically, because all combustion processes emit these compounds to a greater or lesser degree. □

S.A. Penkett is in the Schools of Environmental and Chemical Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Palaeoanthropology

One source not many

Eric Delson

THE origin of anatomically modern humans has long been a problem in human evolutionary studies. But new results in both molecular biology and palaeoanthropology have made it a hot topic for debate. In a paper in last week's issue of *Science*¹, C.B. Stringer and P. Andrews examine the contributions of both approaches to evaluate the two most widely argued models of recent human ancestry, as well as the systematic position of modern humans and our closest extinct relatives.

The argument centres on whether anatomically modern humans evolved independently in different areas of the Old World or in one particular region. The former view, the regional continuity model, is based on the assumption that some morphological characters have persisted independently in four or five regions for hundreds of thousands of years, with continuity maintained by gene flow and locally important (but poorly understood) selective pressures. Proponents of the opposing single-origin model have championed various source areas over the past century, but most now suggest sub-saharan Africa as the most likely homeland. In this 'out of Africa' hypothesis, the anatomically modern ancestors of all non-African peoples migrated into Eurasia around 100,000 years ago, replacing the established 'archaic *Homo sapiens*' occupants perhaps as a consequence of more 'advanced' technology or physique.

Stringer and Andrews formalize these two models in terms of their theoretical predictions about the human fossil record of the past 300,000 years. In all seven comparisons they tested, the out of Africa model agreed far more closely with current data than the regional continuity hypothesis. For example, if anatomically modern humans evolved in Africa from a population of archaic *H. sapiens*, the oldest fossils with modern features would be expected to occur in or near Africa, those found in Europe and eastern Asia being rather younger. On the other hand, if modern humans evolved independently in several regions, there should be no clear temporal pattern to their occurrence (or perhaps contemporaneity might be expected worldwide). The former pattern is in fact observed.

The earliest known anatomically modern humans come from eastern and southern Africa and are about 100,000 years old. The recent report (see ref. 2) of a 92,000-year-old anatomically modern population from Qafzeh in Israel, combined with a comparable date for another

nearly modern population at Djebel Irhoud in Morocco³, strengthens the case further. The southern African fossils are thought to show African regional features, while the North African and near-Eastern samples could represent elements of an ancestral Eurasian population at the point of entering the Northern Hemisphere. Modern humans are unknown from western Europe or Australia before about 35,000 years ago, when the fossils already have at least some regional characteristics



This comparison of (left to right) a Neanderthal from La Chapelle-aux-Saints, an early modern *H. sapiens* from Cro Magnon, and another Neanderthal from La Ferrassie illustrates the problem of the origin of modern *H. sapiens* in Europe. These fossils all derive from caves or rock shelters in France, and the Neanderthal specimens may be only 10,000–20,000 years older than the Cro-Magnon (dated at about 30,000 years ago). Did the Cro-Magnon completely replace the Neanderthals, or was there substantial contact and hybridization between them during the period 40,000 to 30,000 years ago? (Courtesy of C.B. Stringer, by permission of Musée de l'Homme, Paris.)

(the interpretation that the archaic appearance of some Australian crania may be due to cultural practices of deformation rather than persistence of *H. erectus* features is gaining ground⁴).

Genetic and molecular data indicate a major dichotomy between most Africans and all other living humans (see ref. 4). Dates derived from such studies suggest a split more than 100,000 years ago, but these have been questioned on several grounds; my own hesitation relates to the documented variations⁵ between the evolutionary rates of mitochondrial and nuclear DNA.

A third line of evidence (one not emphasized by Stringer and Andrews) is archaeological. The apparent similarity between some Neanderthal artefact assemblages and those associated with early anatomically modern humans such as at Qafzeh and Irhoud belies their anatomical distinction. Archaeological assemblages from southern Africa (for example, the Howieson's Poort and perhaps Pietersburg variants of the MSA) and the southeastern Mediterranean (for example, pre-Aurignacian) which may be more than 90,000 years old include arte-

facts strongly reminiscent of Late Palaeolithic forms but cannot yet be unequivocally associated with human fossils. Perhaps the delay from 90,000 years ago to when anatomically modern humans dominated the Old World was due not only to the time it took to dislodge the successfully adapted Neanderthals and their contemporaries², but also to the time required for the full development of Late Palaeolithic technology and its spread throughout modern *H. sapiens* populations.

Finally, Stringer and Andrews restrict usage of the name *H. sapiens* to anatomically modern populations only, as opposed to a wider usage encompassing Neanderthals and other archaic forms as subspecies. If one or more species of middle

Pleistocene *Homo* is recognized between *H. erectus* and *H. sapiens*, this emphasizes the unique origins and distinctions of the latter. On the other hand, it also clouds the apparent continuity of all varieties of *Homo* in the later Pleistocene and may lead to recognition of greater taxonomic distinction among living humans than is warranted. Study of the morphological diversity within modern populations by comparison with intragroup variation of Neanderthals and other archaic humans is now needed to determine whether these populations are best placed in separate species or retained within a temporally and geographically polytypic *H. sapiens*, as a valuable legacy of the regional continuity model. □

1. Stringer, C.B. & Andrews, P. *Science* **239**, 1263–1265 (1988).
2. Stringer, C.B. *Nature* **331**, 565 (1988).
3. Hublin, J.J. in *Proc. II Int. Congr. hum. Paleont.* (ed. Giacobini, G.) (Jaca, Milan, in the press).
4. Wainscoat, J. *Nature* **325**, 11 (1987).
5. Vawter, L. & Brown, W.E. *Science* **234**, 194–196 (1986).

Eric Delson is Professor of Anthropology at Lehman College and the Graduate School, City University of New York and is in the Department of Vertebrate Paleontology at the American Museum of Natural History, New York, New York 10024, USA.

Amorphous silicates

Easy transformations in glasses

Raymond Jeanloz

TETRAHEDRALLY coordinated SiO_4 is the primary building block of silicate melts and glasses at the Earth's surface, but what about at depth? Does the silicon coordination of melts change with pressure? These are significant questions because melting is one of the primary mechanisms by which the planetary interior evolves; the atomic packing that largely determines the properties of silicate melts ultimately governs much of the Earth's internal dynamics. Edward M. Stolper and Thomas J. Ahrens have recently proposed¹ a model in which silicate melts and glasses transform continuously from tetrahedral to octahedral (SiO_6) coordination under pressure. That such a transformation occurs has previously been inferred from the modelling of high-temperature shock-wave measurements on silicate melts: density increases so rapidly with pressure that it is most easily explained in terms of an increase in coordination². What is new about this model is the idea that the change in atomic packing could be continuous and reversible, even at room temperature. This is surprising as one might expect that a major change in coordination would require bonds to be broken, an energetically costly process for the strong Si–O bond.

Stolper and Ahrens propose a displacive mechanism involving an accordion-like compression of the chains and rings of tetrahedra that make up the structures of amorphous silicates (Fig. 1). The corner-linked SiO_4 tetrahedra are made to rotate toward each other to the point that only minor distortions are required to rearrange the oxygen ions into octahedral configurations around each silicon. The main motivation for such a proposal derives from our recent spectroscopic observations of pressure-induced coordination changes occurring in silicate glasses at room temperature³. Remarkably, the transformations are entirely reversible, although they proceed over a broad pressure range (several tens of gigapascals, depending on the glass composition). Moreover, the large increase in density observed in high-temperature samples subjected to shock waves takes place over the same pressure interval, thus confirming that the activation energy involved is negligible. In short, the data indicate a displacive rather than a thermally activated mechanism.

The new model of Stolper and Ahrens is entirely geometrical, emphasizing the survival of undistorted tetrahedra to large compressions. To make the point that octahedral and tetrahedral coordinations of oxygen around silicon can be viewed as

geometrically closely related, Stolper and Ahrens show how the structure of stishovite, the high-pressure crystalline polymorph of SiO_2 , can be described in terms of irregular rings of distorted tetrahedra (Fig. 2). Although surprising, this does not mean that the transformation of crystalline silicates from tetrahedral to octahedral coordination is displacive. Experiments clearly exclude this possibility^{4,5}.

A key point in Stolper and Ahrens' model is that the tetrahedra are allowed to distort in the last step of the coordination change, once Si–O–Si bond angles of 100° are achieved. That such distortion is required is evident from the locations of the silicon ions, most of which must move through edges or faces of the tetrahedra as the octahedra are formed (Fig. 1c and d). It is precisely this step that seems to require a large activation energy in crystalline structures, yet occurs at essentially no

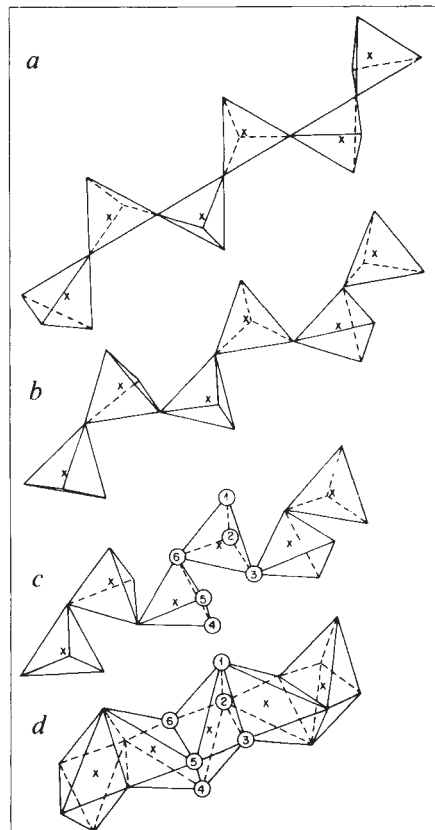


Fig. 1 Stolper and Ahrens' model for the continuous transformation of tetrahedral SiO_4 to octahedral SiO_6 . a–c, The reduction of the Si–O–Si bond angle from 180° to 140° to 100° . d, The new octahedral coordination with the oxygen ions located in the same positions as in c. Oxygen ions are at the vertices (numbers show equivalent ions); crosses, silicon ions. (From ref. 1.)

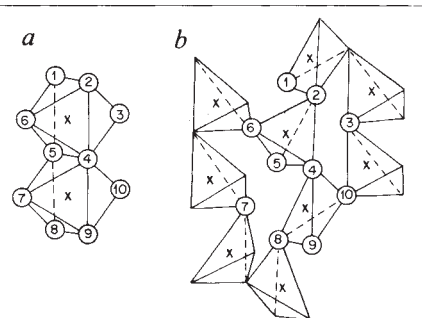


Fig. 2 Description of oxygen packing in stishovite: a, edge-sharing octahedra, usual view; b, in terms of distorted tetrahedra. Oxygen ions at vertices (numbers show equivalent ions); crosses, silicon ions. (From ref. 1.)

energetic cost in the less constrained environment of silicate glass and melt structures. How this might be is still unclear, but it is interesting that previous measurements of crystallization from glass provide independent support for this aspect of the model (M. Aziz, personal communication). Specifically, the rate at which quartz grows in SiO_2 glass is strongly enhanced with pressure, and the activation energy for this bond rearrangement process is expected to vanish above 10–15 gigapascals (ref. 6). Distortion of the tetrahedra towards an octahedral configuration could thus begin at much smaller compressions than invoked by Stolper and Ahrens^{3,7}. This has the advantage that Si–O–Si bond angles need not become smaller than about 120° , agreeing with empirical and theoretical results indicating very high energies for smaller angles^{8,9}.

The main point of Stolper and Ahrens' article is that coordination changes in amorphous silicates, even in solid glasses, occur much more easily under pressure than in the corresponding crystals. A displacive rather than a thermally activated mechanism seems to be involved in transforming the silicon from a tetrahedral to an octahedral environment. Thus, it seems inevitable that the physical properties of silicate melts deep inside the Earth are strongly influenced by these surprisingly easy coordination changes¹⁰. □

1. Stolper, E.M. & Ahrens, T.J. *Geophys. Res. Lett.* **14**, 1231–1233 (1987).
2. Rigden, S.M., Ahrens, T.J. & Stolper, E.M. *J. geophys. Res.* **93**, 367–382 (1988).
3. Williams, Q. & Jeanloz, R. *Science* **239**, 902–905 (1988).
4. Yagi, T. & Akimoto, S. *Tectonophysics* **35**, 259 (1976).
5. Knittle, E. & Jeanloz, R. in *High-Pressure Research in Mineral Physics* (eds Manghnani, M.H. & Syono, Y.) 243–250 (Am. Geophys. Un., Washington DC, 1987).
6. Fratello, V.J., Hays, J.F., Spacen, F. & Turnbull, D. *J. appl. Phys.* **51**, 6160–6164 (1980).
7. Meade, C. & Jeanloz, R. *Phys. Rev. B* **35**, 236–244 (1987).
8. O'Keeffe, M. & Hyde, B.G. *Acta Cryst.* **B34**, 27–32 (1978).
9. Gibbs, G.V. *Am. Mineral.* **67**, 421–450 (1982).
10. Angell, C.A. *et al. Science* **218**, 885–887 (1982).

Raymond Jeanloz is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, on leave from the Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA.

DNA repair

Views of unity and diversity

C.S. Downes

DNA is continually assaulted by undesirable reactions, and repair systems correct the resulting damage. Imperfect correction produces mutations, which could be undesirable for the individual, and in the long term may contribute to evolution. The rates of DNA sequence evolution vary considerably between taxonomic groups¹, as do the rates of onset of individual phenomena, such as tumours or ageing, which can be more or less plausibly attributed to failures in DNA repair. It is not certain how far these variations between taxa are affected by different responses to DNA damage. At a recent meeting* both the unity and the diversity of DNA repair were discussed.

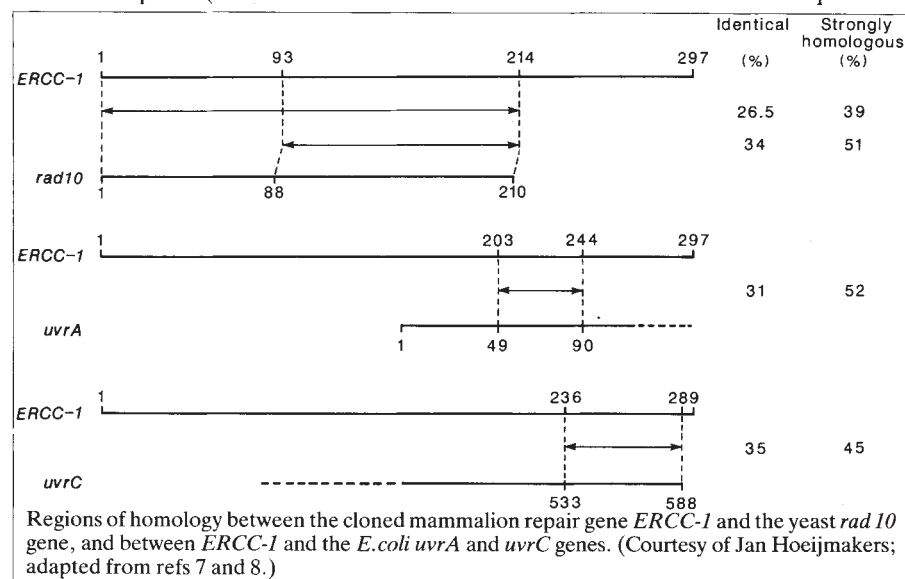
The essential unity of DNA repair is particularly attractive to those who study repair in bacteria, or in what were once called lower, but are now as a gesture of anti-phyllism redefined as smaller, eukaryotes. Repair in *Escherichia coli* is understood in some detail; repair mutants are available in yeast and *Drosophila* in a profusion not attained for any mammalian species. Even in *E. coli*, the repair systems are diverse, and different systems have overlapping functions. R. H. Haynes (York University, Toronto) identified this complexity as an illustration of an axiom in information theory, Dancoff's principle of maximum error, according to which tolerable fidelity can be achieved with equipment of poor precision only by incorporating repeated and overlapping testing procedures. According to this view, a multiplicity of repair systems is always necessary, but the elements of the systems can be fundamentally similar in different taxa. The hope is that the function of DNA repair genes in the less amenable larger eukaryotes will be made clear through 'evolutionary walking', using similarities of amino-acid sequence between gene products from different phyla as indicators of a common mechanism.

Indeed, the only mammalian repair gene that has been cloned and sequenced, the *ERCC-1* human gene, which restores excision repair to one class of Chinese hamster repair-defective mutants², was described by J. H. J. Hoeijmakers (Erasmus University, Rotterdam) as having considerable similarity in one part of its sequence with the *E. coli uvrA* and *uvrC* genes, and in another with the yeast *rad10* gene (see figure); all three of these genes act in excision repair. Functions for the products of the bacterial genes were

described by L. Grossman (Johns Hopkins University, Baltimore): the *uvrA* product recognizes large distortions in the DNA helix and binds near them, starting the formation of a repair complex in which the *uvrC* product is needed to make strand breaks around the damaged region. How the *rad10* product works is less clear.

Sequences of several more mammalian repair genes should soon be available for evolutionary walking, especially from L.H. Thompson (Lawrence Livermore

is not, however, universal. Further work at Stanford indicates that the removal of alkylation damage is not specific for transcribed regions; and O. Kessler (Technicon, Haifa) finds that activation of genes during rat myogenesis does not result in enhanced excision of ultraviolet damage sites therein. The mechanism for this differential repair, when it exists, could involve attachment of the repair sites to the nuclear matrix, which is also the site of transcription. L. Mullenders (University of Leiden) demonstrated that in cells from patients with Cockayne's syndrome, an ultraviolet-sensitive disease with no overall defect in ultraviolet repair, transcribed genes are not selectively repaired and repair sites are not, as in normal cells, enriched in matrix-associated sequences.



Laboratories, New Mexico). Even without this information, extrapolation from smaller eukaryotes may yield insights; the demonstration by S. Jentsch (MIT) that the yeast *rad6* repair gene encodes a histone ubiquitin-conjugating enzyme³ strongly implies that ubiquitin is involved elsewhere in repair, an unexpected prediction.

The other theme at the meeting was that repair systems can vary very significantly between different orders, let alone phyla. Again, this theme was most visible in ultraviolet excision repair. P. Hanawalt and collaborators (Stanford University), studying the loss of ultraviolet-induced pyrimidine cyclobutane dimers, show that there are differences between hamster cells, which concentrate their excision repair on active genes, and human cells, which are less discriminatory⁴. I. Mellon (Stanford University) has further refined the distinction between hamster and human excision. In the former, repair concentrates on the transcribed strand of the active gene, and repair elsewhere is never complete⁵; in the latter, repair outside the transcribed strand is slower but ultimately complete. This generalization

The difference between excision repair strategy in hamsters and humans is annoying enough; it can probably be extended to a difference between rodents and primates, which at least *in vitro* have respectively low and high capacities for excision repair (and incidentally have contrastingly short and long lifetimes, and high and low rates of genome evolution). There may be even less unity. D. Mitchell (University of Texas, Smithville) demonstrated that one hamster ultraviolet-sensitive mutant line cannot excise cyclobutane dimers, but can repair the minor photoproduct (6-4)di-pyrimidine normally, whereas a human ultraviolet-sensitive line deficient in excision of both lesions is restored to normality when it regains the capacity to excise the (6-4)photoproduct, but not cyclobutane dimers. One hopes this should not be generalized to mean that hamsters and humans are preferentially sensitive to different DNA lesions.

Furthermore, it has been known for some time that human cells from ultraviolet-sensitive patients suffering from xeroderma pigmentosum, and deficient in excision repair, can be separated into several genetic complementation groups,

* The UCLA symposium *Mechanism and processing of DNA damage* Taos, New Mexico, 24-30 January 1988.

currently nine in number. It is worrying that these groups have not yet been found to correspond to the seven complementation groups of excision-defective hamster mutants, although human genes can restore hamster repair *in vitro*. The nature of a mammalian repair-complementation group had itself been thought to be clear; each was supposed to represent a different gene. There is now some concern about this. A. R. Lehmann (University of Sussex) described the complexities of trichothiodystrophy, another ultraviolet-sensitive disease in which the symptoms are very different from xeroderma, but in which cells from some patients selectively fail to complement xeroderma group D.

An ingenious explanation of some difficulties in understanding complementation groups was advanced by W.C. Lambert (New Jersey Medical School, Newark), who proposes that many human DNA repair diseases show co-recessive inheritance⁶. In his model, defective phenotypes would be expressed only in individuals homo- or hemizygous for defective alleles at more than one locus. This would cause a lack of complementation between two syndromes sharing one co-recessive locus; it would also diminish the complexity of excision repair, because a few co-recessive genes can generate a larger number of complementation groups. The model predicts a high frequency of carriers for co-recessive repair mutations; little dependence of clinical xeroderma on first-cousin marriages (this is being tested); and great difficulties in correcting repair defects by human gene transfection. Indeed, xerodermas turn out to be singularly resistant to such correction; but this could be a consequence of limited DNA uptake. Partial correction of xeroderma group A cells by transfected mouse DNA was announced by K. Tanaka (Osaka University), who showed what can be done by 'macro-experiments' in which one repair-positive colony occurs in 3,000 dishes.

The study of excision repair is only secondarily relevant to mutation and evolution, as errors arise largely in cases where damage is not removed, but damaged DNA is used as a template for replication (as must happen to a greater extent in hamsters than in humans). But the error-prone processing of such damage has been harder to investigate. In *E. coli*, the damage-inducible, error-prone SOS system is becoming understood in great detail. Error-prone processing in larger eukaryotes will probably be elucidated by the use of shuttle vectors that can be mutagenized in a mammalian cell and transferred to bacteria for analysis. Among the virtuoso feats attained with such systems, perhaps the most promising is the work presented by K. Dixon (NIH, Bethesda), using a simian virus 40-based shuttle vector for studies of

mutagenesis both in mammalian intact cells and *in vitro* during replication by cell extracts.

How far mammalian repair, error-prone or otherwise, is inducible by damage is an unanswered question. If the induction of human DNA repair is ever understood, there may conceivably be remarkable prophylactic opportunities. That some damage responses are highly inducible was demonstrated for human cells by P. Herrlich (Kernforschungszentrum, Karlsruhe). Some of the human elements stimulated have been identified, mostly as general stress-response genes, and the sequences responsible for their transcriptional control studied; ultraviolet damage to DNA, or cell-membrane actions of serum or the tumour promoter TPA, seem to stimulate transcription at the same regulatory sites but by different pathways. A different set of ultraviolet-inducible genes has been identified in hamster cells by A. J. Fornace (NIH, Bethesda), using

hybridization subtraction to remove non-induced elements from a complementary DNA library prepared from irradiated cells. At least 35 different induced genes have been identified: they are not inducible by TPA, many are also inducible by other DNA damaging agents, and some are over- or under-expressed in ultraviolet-sensitive mutants. Fortunately for the unity of repair, many are also expressed in human cells. □

1. Britten, R. J. *Science* **231**, 1393-1398 (1986).
2. Van Duin, M. *et al. Cell* **44**, 913-923 (1986).
3. Jentsch, S., McGrath, J. P. & Varshavsky, A. *Nature* **329**, 131-134 (1987).
4. Bohr, V. A., Philips, D. H. & Hanawalt, P. C. *Cancer Res.* **47**, 6426-6436 (1987).
5. Mellon, I., Spivak, G. & Hanawalt, P. C. *Cell* **51**, 241-249 (1987).
6. Lambert, W. C. & Lambert, M. W. *Mutat. Res.* **145**, 227-234 (1985).
7. Hoeijmakers, J. H. J. *et al. Cold Spring Harb. Symp. quant. Biol.* **51**, 91-101 (1986).
8. Doolittle, R. F. *et al. Nature* **323**, 451-453 (1986).

C.S. Downes is in the CRC Mammalian Cell DNA Repair Group, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK.

Mineralogy

Lamprophyres a source of gold

D. Wyman and R. Kerrich

THE origin of Archaean-type gold and silver (Au, Ag) vein deposits is highly controversial, primarily because of the complexity of their geological setting. Rich gold lodes were deposited in brittle-ductile faults in ancient Archaean greenstone belts and some younger tectonic belts like the cordilleran belt in North America. The faults cross-cut volcanic and intrusive igneous rocks of diverse composition, including lamprophyres, as well as clastic and hydrothermal sediments. Indeed, gold deposits occur within the entire range of rock types present. A recurrent theme in many hypotheses of the genesis of (Au, Ag) vein deposits is that there is a special source rock, characterized by an intrinsically enhanced gold content. Rock and Groves, on page 253 of this issue¹, argue that the spatial and temporal association of (Au, Ag) vein deposits with lamprophyres is a direct consequence of the derivation of these igneous rocks from a mantle source region characterized by primary gold enrichment. An alternative interpretation² is that gold mineralization and lamprophyres are generated in a common tectonic setting, but otherwise are the products of distinct geological processes.

Rock and Groves clearly favour the first hypothesis and outline evidence to indicate that all lamprophyre types are gold-rich. They then suggest that calc-alkaline lamprophyres are related to gold mineralization because they are the only member of the clan of lamprophyres that undergo significant interaction with the crust, a

process that may allow for release of gold into crustal volatile systems. Two questions arise: first, is the anomalous gold content of lamprophyres primary or secondary; and second, what is the significance of the association between (Au, Ag) vein deposits and lamprophyres?

The background, or primary gold content of most rocks is in the range 0.5-2 parts per billion (ref. 3). To account for all the gold present in the rich lodes of a mining camp such as Timmins, Canada, where more than 3,000 tonnes of gold have been mined, hundreds of cubic kilometres of source rock must have been involved if the concentration process is about 10 per cent efficient⁴. The appeal of a rock type intrinsically enriched in gold, therefore, is that the source volume could be correspondingly smaller.

The accurate measurement of the primary gold content of any rock type is a subtle problem, particularly if the rock is near a gold deposit. If the gold content is anomalously high, is it an inherent, primary chemical feature of a rock, or does it result from contamination by the gold mineralizing system? Anomalous metal abundances are ubiquitous in dispersion halos around ore deposits, but may not be revealed by optical examination alone. A possible resolution relies on comparing the content of platinum-group metals with gold content. Gold is thought to behave coherently with the platinum-group metals during some mantle-melting processes, but in crustal aqueous hydrothermal systems, iridium and platinum

are less mobile than gold, providing a benchmark against which gold enrichment or depletion can be measured⁵.

In a view that avoids the hypothesis of gold-enriched source rocks, it is the nature of large-scale hydrothermal leaching that is selective with respect to gold and any rock is an adequate source⁶⁻⁸. It is possible that gold-enriched rocks such as lamprophyres coexist with gold-mineralizing systems, but that the gold in the deposits is derived from an independent source, and the relationship between gold deposits and lamprophyres reflects a tectonic control.

The highest gold contents in Rock and Groves' data¹ come from variably deformed and hydrothermally altered calc-alkaline lamprophyres from the Canadian Arrow gold mine. Such occurrences of gold-enriched lamprophyres in gold mines are generally interpreted as the product of the late addition of gold, especially because, in most instances, the dykes are clearly not the locus of mineralization, but could have acted as dams or barriers to gold-bearing fluids⁹. There is additional evidence for their model in the elevated gold content of lamprophyres from south-west Scotland which, apparently, do not lie within the dispersion halos of large-scale gold deposits. The evidence is, however, open to at least one other interpretation. Whereas Rock and Groves attribute the low gold content of many calc-alkaline lamprophyres to either gold loss during crustal interaction or the non-uniformity of mantle source regions, it could be that apparently unaltered gold-rich dyke magmas have been contaminated during their ascent. Vein-type gold mineralization occurred in the same region as the Scottish dykes¹, and thus could have been a source of secondary gold enrichment in the lamprophyres.

Gold-bearing vein systems and lamprophyres generally developed along the same structural discontinuities, or fault systems, which guided their emplacement^{1,2}. Therefore, both geological events are constrained to the same relatively small volumes, at least in the upper crust. The close spatial and temporal association of calc-alkaline (or shoshonitic) lamprophyres and mesothermal (deep) gold-mineralization occurs in several gold camps of various ages dating back to the Archaean^{1,2,9,10}. Early prospectors in the Canadian Shield knew this 50 years ago, but the nature of the functional relationship remains unresolved. The current interest in lamprophyres stems from their specific temporal and tectonic setting, which may help elucidate contemporaneous tectonomagmatic processes at deeper crustal and mantle depths and, perhaps, the gold-mineralization system itself.

Rock fragments (xenoliths) entrained in rising magma provide evidence that the

dykes have sampled a wide range of crustal material in these structural zones. In the contamination model, the high gold content in lamprophyres is largely accidental and gold abundance can vary markedly within a dyke swarm over comparatively short distances, or between lamprophyre dyke swarms in different gold provinces. Hence, the occurrence of barren lamprophyres near gold deposits of similar age can be attributed to the lack of intersection of individual dykes with gold-rich zones², or to the loss of primary gold from the lamprophyre magma at an earlier stage¹.

Significant lode-gold deposits of Archaean greenstone belts can be generated before, after and with regional metamorphism. On the other hand, geological mapping shows that lamprophyres are emplaced exclusively in the late-tectonic regime dominated by obliquely compressional tectonics that follow severe deformation and metamorphism². If these observations stand up to examination, then either lamprophyres are a non-unique component of gold deposits, or lamprophyre magmas are generally present, but some factor in the tectonics of the premetamorphic deposits discriminates against their emplacement. Similarly, lamprophyres are not generally associated with the large recent (Au, Ag) vein deposits of Nevada, so that there could be an alternative mineralizing process (or several) or lamprophyres could be involved but remain unexposed.

Lamprophyres and (Au, Ag) vein deposits are characterized by abundant carbonate minerals, the stable-isotope composition of which could identify a genetic relationship between the two¹⁰. If calc-alkaline lamprophyres consist of crustally contaminated gold-bearing lamproites, as suggested by Rock and Groves¹, then further analyses should show that the lamproites have even higher average gold content. A marked and consistent difference in the volatile content and volatile composition of gold-rich and gold-poor calc-alkaline lamprophyres can also be predicted by Rock and Groves' model. □

1. Rock, N.M.S. & Groves, D.I. *Nature* **332**, 253-255 (1988).
2. Wyman, D. & Kerrich, R. *Econ. Geol.* (in the press).
3. Tilling, R.I., Gottfried, D. & Rowe, J.J. *Econ. Geol.* **68**, 168-186 (1973).
4. Kerrich, R. *Can. Inst. Min. Metall. spec. Pap.* **27**, 1-79 (1983).
5. Keays, R.R. in *Gold '82* (ed. Foster, R.P.) 17-52 (Balkema, Rotterdam, 1984).
6. Kerrich, R., Fryer, B.J., King, R.W., Willmore, L.M. & Van Hees, E. *Contr. Miner. Petrol.* **97**, 156-168 (1987).
7. Groves, D.I. & Phillips, G.N. *Ore Geol. Rev.* **2**, 287-322 (1987).
8. Nesbitt, B.E., Murowchick, J.B. & Muehlenbachs, K. *Geology* **14**, 506-509 (1988).
9. Gilbert, G. in *Can. Inst. Min. Metall. Jubilee Vol.* 189-196 (1948).
10. O'Neil, A.M. & Kerrich, R. *Can. J. Earth Sci.* **23**, 324-343 (1986).

D. Wyman and R. Kerrich are in the Department of Geological Sciences, University of Saskatchewan, Saskatoon, S7N 0W0, Canada.

Daedalus

Arrest the offenders!

ILLEGAL car-parking is a statistical game played between the car-owning public and the authorities. A potential illegal parker will risk quite heavy penalties if he perceives the chance of being caught as small enough. In several countries the authorities now impose a very severe punishment indeed: the wheel-clamp. The offending car is immovably shackled, and its hapless owner faces much trouble and expense before he can release it.

But even this deterrent is ineffective if the chance of being clamped is too small. So Daedalus now proposes a scheme to make it 100 per cent. It is based on those anaerobic glues which set solid in the absence of oxygen, and which are used by engineers to lock bolts into deep holes.

DREADCO's chemists are devising a thick, resilient, slightly tacky version of such a glue, for use as a road-surface. It does no harm to law-fully-moving traffic; indeed its slight tack enhances road-grip and inhibits skidding. But an illegally parked car cuts off the air from the glue beneath its tyres. The glue sets, and the returning owner finds his car glued to the spot. This elegant technology makes the crime not only match, but actually cause the punishment!

How will a glued car get free again? The DREADCO chemists hope to make their glue strong enough to strip the tread from the tyres when the owner, raging, finally drives away. He will have to drive home slowly and carefully, his pitted tyres advertising his disgrace, and will have to buy a whole new set. A special crew with blow-torches and scrapers may have to travel the roads removing the tread-sections left behind by glued vehicles, otherwise later cars with the same wheelbase might manage to park exactly on them, neatly evading the penalty. But Daedalus hopes to make his anaerobic glue reversible. Once a tread-section has been detached from a tyre, oxygen should slowly diffuse through the torn surface to release the section.

Many refinements of this cunning scheme are possible. Fifteen-, thirty-, and sixty-minute glues could be compounded, making possible an automatic and pitiless parking-meter system. Very fast-acting glues could be spread on those areas of road that you should not stop on even briefly. And areas of pavement vulnerable to street offences could also be treated. Hawkers, pamphleteers, prostitutes and market researchers would be forced to keep moving, or risk losing their shoes. The old English offence of 'loitering with intent' would be made very difficult; and offenders, glued to the spot, would be easy to arrest.

David Jones

Magnetic reversal rate and sea level

SIR—A negative correlation between the long-term eustatic sea level and the geomagnetic reversal rate for the past 150 million years (Myr) was recently analysed by Gaffin¹. We had previously pointed out the existence of such a correlation^{2,3}, although we did not do a complete mathematical analysis; nor did we detect the relatively small 10-Myr phase shift reported by Gaffin. He attributes the correlation to long-term coupling between the flow pattern in the core (possibly responsible for spontaneous geomagnetic reversals) and the configuration of the mantle (which determines oceanic ridge volume and possibly sea level).

Our model presents a different explanation. We attribute a substantial fraction of the geomagnetic reversals to sudden changes in the moment of inertia of the Earth that occur during abrupt sea-level drops, when water is transferred from the oceans to continental ice at high latitudes; the faster rotation of the crust and mantle creates a velocity shear in the liquid core which disassembles the dynamo. A definite prediction of our theory is that the individual reversals (or geomagnetic excursions) should not lag the individual sea-level drops by more than a few times the characteristic convection time in the liquid core, a few thousand years. Our model successfully accounts for the initial sharp drop in the magnitude of the field at the beginning of a reversal and extended period of low field dominated by high-order multipole moments^{4,5}, and also several previously unexplained correla-

tions, including the association of geomagnetic reversals with microtektite layers⁶, craters⁷, biological extinctions⁸ and short periods of cold climate⁹. The cooling that causes the abrupt sea-level drop could be induced by extraterrestrial impacts, volcanic eruptions or anything else that suddenly alters the global energy balance. Thus external events can change the flow pattern in the core, and induce geomagnetic reversals.

The negative correlation between the long-term eustatic sea level and the average rate of reversals is explained in our paper by the warmer Earth climate that existed during the period when the sea level was high, and the fact that larger (hence less frequent) cooling events were required to build up continental ice. A change in the rate of these events with time could give a phase shift such as that reported by Gaffin¹, although the individual reversals would continue to be coincident in time with individual cooling events.

R.A. MULLER

D.E. MORRIS

Lawrence Berkeley Laboratory,
University of California,
Berkeley, California 94720, USA

1. Gaffin, S. *Nature* **329**, 816–819 (1987).
2. Muller, R.A. & Morris D.E. *Geophys. Res. Lett.* **13**, 1177–1180 (1986).
3. Schwarzschild, B. *Physics Today* **40**, 17–20 (February 1987).
4. Jacobs, J.A. *Reversals of the Earth's Magnetic Field*, Ch.3 (Hilger, Bristol, 1984).
5. Valet, J.P., Lal, C. & Turcholka, P. *Nature* **322**, 27 (1986).
6. Glass, B.P., Swincki, M.B. & Zwart, P.A. *Proc. lunar planet. Sci. Conf.* **10**, 2535–2545 (1979).
7. Pohl, J. *Geologica bav.* **75**, 329–348 (1977).
8. Keany, J. & Kennett, J.P. *Deep Sea Res.* **19**, 529 (1972).
9. Krishnamurthy, R.V. *et al. Nature* **323**, 150–152 (1986).

Punctuated equilibrium prevails

SIR—Sheldon's¹ study of patterns of morphological change within eight lineages of mid-Ordovician trilobites provides additional evidence that palaeontological data are indeed relevant to contemporary evolutionary theory. Yet Maynard Smith's² accompanying comment reveals an inaccurate grasp of the implications of Sheldon's study and the evolutionary issues that surround the idea of 'punctuated equilibria'. There is still a gap between the perception of evolutionary pattern and the efforts of theorists to understand causal relationships underlying evolutionary history.

Sheldon has studied changes in pygidial pleural rib counts in eight separate species-lineages of trilobites from the Builth inlier, spanning an interval of about 3 million years (Myr). With careful sampling, he has admirably taken advantage of the main strengths of the marine invertebrate fossil record. He has demonstrated within-sample variation in pygidial pleural rib counts, and has shown that modes may shift through time. We dispute his claim

for eight parallel lineages of gradually increasing ribbiness. At least one lineage zig-zags and ends up where it started; two others show two periods of stability separated by a jump with no intermediates — but Sheldon's chart joins the two segments with a diagonal line, implying gradual transition when no evidence exists.

Both Sheldon and Maynard Smith contrast Sheldon's results with 'stasis' — an empirical cornerstone of our model of punctuated equilibria³. The implication seems to be that stasis necessarily entails no within-population, or even geographically based among-population, variation; such a caricature of our conception of stasis further implies a rigidly monolithic straight-jacketed lack of variation through the entire duration of a species-lineage. Yet we have always acknowledged that such variation exists at any one time in a species' history. The failure to convert this variation into substantial anagenetic change is central to punctuated equilibria.

To quibble whether Sheldon's examples fall within 'phyletic gradualism' or 'stasis'

is to miss the larger point. Gingerich⁴ has presented patterns of size increase and decrease of various Eocene mammalian taxa forcefully reminiscent of Sheldon's trilobite patterns. Lande⁵ analysed some of Gingerich's data, concluding that the amount of change accrued over several million years of Eocene time was almost incomprehensibly negligible — far too slight to be distinguished from the effects of genetic drift, and too slight, perhaps, even to be comfortably explained as the consequence of natural selection. Most evolutionary biologists (including ourselves) see natural selection as a far more efficient causal agent of evolutionary change than Lande was able to demonstrate.

Much the same can be said of Sheldon's trilobites. A net increase of two or three pleural pygidial ribs over 3 Myr in these trilobite lineages can hardly address the issues of patterns, and underlying processes, of changes that mark closely related, stratigraphically overlapping yet geographically disjunct congeneric species; and such small changes cannot have any real bearing on the origin of morphological differences that characterize taxa above the species level. The interplay of stochastic and deterministic processes tinkering with pygidial pleural rib numbers on the Builth trilobites for 3 Myr yields little insight on the general dynamics of trilobite diversification. When taxa above the species level appear in the trilobite record, they do so rather abruptly. Evolutionary theorists must acknowledge^{6,7} that simple extrapolation of the amount of change that Sheldon documents over 3 Myr would require vastly more than the 340-Myr recorded span of trilobite history to yield the morphological diversity attained within the Ordovician alone. Such rates of evolution are simply too slow to account for observed patterns of diversification.

Sheldon and Maynard Smith both think that the evidence of 'punctuation' resides in sudden, even quantal, morphological shifts within a stratigraphic succession of fossils from a single lineage. In contrast, we have always demanded stratigraphic overlap between two distinct taxa interpreted to be sister species, with the presumed ancestor persisting for some time alongside its putative descendant. The problem is to understand the origin of the consistent differences between the two species; only in such a context can (allopatric) speciation models be applied — the very heart and soul of punctuated equilibria.

Sheldon, citing Hughes's diagnoses^{8–10} of several non-stratigraphically overlapping species (chronospecies), and further citing his demonstration that Hughes's taxa are not discrete, but rather intergrade, concludes such a use of species names falsely gives the impression of punctuated

equilibria. But of course we recognize that a naming system based on discrete entities will make any sequence look like a discontinuous array, and the debate has properly focused on the statistics of measured morphological change, not on published compendia of names. The evidence for punctuated equilibria is morphological, not nomenclatural.

Once again, we are confronted with profoundly different ontological conceptualizations of what species are: Sheldon, following traditional palaeontological practice, sees a species as arbitrarily delineated segments of an evolving continuum. As Mayr¹¹ has said, a relatively complete fossil record would prohibit facile discrimination of species so conceived. In contrast, we follow the other ontological thread Mayr¹¹ sets out, seeing species as discrete reproductive communities. Like Simpson¹² and Wiley¹³, we have no trouble imagining such reproductive communities persisting for considerable segments of geological time as lineages. The resultant picture of species as spatiotemporally discrete and independent reproductive lineages derives from punctuated equilibria, and is fraught with further implications for evolutionary theory — including the debate over 'species selection'.

Maynard Smith² goes far beyond the measured tones of Sheldon's report. He cites species selection as his principal evidence that punctuated equilibria is just another manifestation of such palaeontological mistrust. He defines species selection as a causal theory of the origin and modification of organismal adaptation. Naturally, he has no trouble dismissing such an absurd postulate, saying that "there never was much sense in the idea anyway". Of course it makes no sense — stated this way. Maynard Smith seems not to realize that this characterization of species selection is entirely of his own making. Where discussions of our own (and colleagues¹⁴⁻¹⁶), on species selection address patterns of organismal phenotypic evolution, our idea is that 'sorting'

of species changes the spectrum of genetically based phenotypes available for the ongoing populating of ecosystems — and for further events in the genealogical history of life.

Morphological adaptations are complex properties of organisms, and therefore presumably evolved directly at the level of organisms. But patterns of waxing and waning of diversity occur at the species level. The myopia of traditional thinking lies in the assumption that such issues of the species level extrapolate directly from adaptation at the organismic and population levels. But we do not believe that the world contains millions of species of beetles and only a handful of pogonophorans simply because beetles represent a successful adaptive design. Propensity for speciation and resistance to extinction must also be considered.

There are other faulty characterizations of punctuated equilibria in Maynard Smith's commentary. He repeats, for example, the canard that palaeontologists invoke only 'developmental constraints' as a source of stasis — rather than 'normalizing selection'. Yet it is palaeontologists¹⁷⁻¹⁹ more than any other cadre of evolutionary biologists who have realized that environmental change results in habitat shifting. Continued persistence with relatively little change will be the rule as long as organisms can continue to recognize suitable habitats — a point only recently appreciated by some non-palaeontological evolutionary theorists^{20,21}.

Palaeontologists don't mistrust population geneticists. Years ago, Henry Fairfield Osborn²², taking exception to the genetically-based evolutionary theories of his day, saw fit to invent his own ("aristogenesis"). G.G. Simpson⁶ put an end to that nonsense, showing that there must be a single, all-encompassing causal theory of evolution — one that integrates the principles of physiological and population genetics with the patterns of actual evolutionary history. We all still seek a more accurate integration. If Darwin gave us a general, very powerful, understanding of the basic mechanism of stasis and modification of organismic adaptations, all evolutionary biologists still await a coherent statement of the context of such stasis and change: a deeper understanding of when, and how much, evolutionary change might be expected to occur. Punctuated equilibria poses the challenging and almost counter-intuitive link between change in the economic adaptations of organisms and the disruption of reproductive communities that is speciation. Some non-palaeontological evolutionary biologists (for example Futuyma²⁰) are beginning to pursue this line of thought. We should realize, though, that we are all darwinians in our focus upon selective processes; but in taking the fundamental change of science seriously, supporters of

punctuated equilibria seek a deeper understanding than that already achieved.

NILES ELDRIDGE

Department of Invertebrates,
American Museum of Natural History,
New York, New York 10024, USA

STEPHEN JAY GOULD

Museum of Comparative Zoology,
Harvard University, Cambridge,
Massachusetts 02138, USA

Energy conversion by gravitational waves

SIR—We have been studying some surprising properties of the paths of particles and light rays in the field of a plane gravitational wave of limited duration — a sandwich wave¹. (Penrose² has pointed out some exotic properties of null cones in this context.) We have shown (manuscript in preparation) that a plane gravitational wave induces large relative accelerations in initially widely separated test particles. Particles which initially are arbitrarily far apart and at relative rest will be constrained to collide within a finite time which is independent of their initial separations.

Here we point out that if such particles are charged, the accelerations will constitute a mechanism for the conversion of gravitational energy into electromagnetic energy: in a plasma, the electrons and the positive charge carriers will be equally affected by the gravitational field. As, however, the electromagnetic radiative reaction will affect the electrons far more than it will affect the positive charge carriers, with their much greater mass/charge ratio, radiation from the positive and negative carriers will not cancel. Professor B. Schutz of University College Cardiff has pointed out that the mechanism of energy transfer cannot be efficient at frequencies below a critical frequency, at which the gravitational wavelength equals the typical interparticle spacing in the plasma. For astrophysical plasmas this critical frequency is in the gigahertz region or higher.

The attenuation of high-frequency gravitational waves in a plasma may therefore be much greater than has hitherto been expected, and the resulting electromagnetic radiation may have distinctive observable characteristics. The details of the energy transfer mechanism, and the applicability of the plane wave approximation near to likely sources of gravitational waves, require further study.

H. BONDI

Churchill College, Cambridge
CB3 0DS, UK

F.A.E. PIRANI

Kings College, London WC2R 2LS, UK

1. Bondi, H., Pirani, F.A.E. & Robinson, I. *Proc. R. Soc. A* **251**, 519–533 (1959).

2. Penrose, R. *Rev. mod. Phys.* **37**, 215–220 (1965).

- Sheldon, P.R. *Nature* **330**, 561–563 (1987).
- Maynard Smith, J. *Nature* **330**, 516 (1987).
- Eldredge, N. & Gould, S.J. in *Models in Paleobiology* (ed. Schopf, T.J.M.) 82–115 (Freeman, San Francisco, 1972).
- Gingerich, P.D. *Nature* **248**, 107–109 (1974).
- Lande, R. *Evolution* **30**, 314–334 (1976).
- Simpson, G.G. *Tempo and Mode in Evolution* (Columbia Univ. Press, New York, 1944).
- Stanley, S.M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 646 (1975).
- Hughes, C.P. *Bull. Br. Mus. nat. Hist., Geol.* **18**, 39 (1969).
- Hughes, C.P. *Bull. Br. Mus. nat. Hist., Geol.* **20**, 115 (1971).
- Hughes, C.P. *Bull. Br. Mus. nat. Hist., Geol.* **32**, 109 (1979).
- Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942).
- Simpson, G.G. *Principles of Animal Taxonomy* (Columbia Univ. Press, New York, 1961).
- Wiley, E.O. *Syst. Zool.* **27**, 17–26 (1978).
- Eldredge, N. *Unfinished Synthesis* (Oxford Univ. Press, New York, 1985).
- Vrba, E.S. & Eldredge, N. *Paleobiology* **10**, 146–171 (1984).
- Vrba, E.S. & Gould, S.J. *Paleobiology* **12**, 217–228 (1986).
- Coope, G.R. *A. Rev. Ecol. Syst.* **10**, 247–267 (1979).
- Eldredge, N. & Gould, S.J. *Evol. Biol.* **7**, 303–308 (1974).
- Vrba, E.S. *J. Afr. J. Sci.* **81**, 229–236 (1985).
- Futuyma, D.J. *Am. Nat.* **130**, 465–473 (1987).
- Lande, R. *Paleobiology* **12**, 343–354 (1986).
- Osborn, H.F. *Am. Nat.* **68**, 193–235 (1934).

Taming fission's bastard

Alvin M. Weinberg

The International Atomic Energy Agency and World Nuclear Order. By Lawrence Scheinman. *Resources for the Future*, 1616 P St, NW, Washington, DC 20036: 1987. Pp.320. Pbk \$16.95 (\$19.95 by post).

ALTHOUGH concern over reactor safety and the disposal of radioactive wastes is largely responsible for today's massive public defection from nuclear energy, among experts, both pro- and anti-nuclear, the connection between nuclear power and nuclear weapons is viewed as fission's ultimate Achilles' heel. Technology can provide safe reactors and waste disposal. No technological fix can prevent a state intent on building a nuclear weapon from doing so: proliferation is, in a sense, an insoluble political problem.

But if nuclear energy is necessary — if for no other reason than to diminish the greenhouse effect — then the issue of proliferation must be resolved. This dilemma brought forth the Acheson-Lilienthal proposal in 1946 to internationalize all nuclear activities under a supreme, worldwide authority. The hope that such gross intrusion on national sovereignty would be acceptable was much too Utopian. Instead, President Eisenhower in 1953 put forward his Atoms-for-Peace proposal: weapons states would contribute fissile material from their stockpiles to a central authority; and an international authority would supervise such Isaiah-like transfer and otherwise encourage peaceful uses of fission. Though Atoms-for-Peace was derided by Robert Oppenheimer, the author of Acheson-Lilienthal, as being "allusive-and-sentimental", it has, in this real world, given rise to the existing international non-proliferation regime.

Thirty-five years after Atoms-for-Peace, the central depository for fissile material remains a vision. On the other hand, the vigorous, and in many ways uniquely effective International Atomic Energy Agency (IAEA) has become the linch-pin of the non-proliferation regime: as Professor Scheinman says, it is "the first attempt of our restless species to combine agreements on arms control with objective and effective verification [of] compliance". Scheinman's scholarly examination of IAEA brilliantly explains how the agency developed over the years, how it now discharges its diverse responsibilities, and how it copes with the inevitable tendencies towards political interference by competing member states.

At present, IAEA has three main functions: broad technical assistance in nuclear science and technology, particularly for the developing world; dissemination

of information on reactor safety, a function that now receives much emphasis in the wake of Chernobyl; and implementation of the safeguards regime, both of the original IAEA protocol and the much more far-reaching protocol of the Non-Proliferation Treaty (NPT). As Scheinman tells us, the first two functions have been executed well, and have caused little controversy. On the other hand, safeguards, which are regarded by the five weapons states, but not by most of the other members, as the central function of IAEA, are a much more controversial and

If one judges IAEA by its fulfilment of the protocols established by its charter, and by the directives of its board of governors, then the agency must be considered an astonishing success. It is generally regarded as the least politicized of the UN-associated bodies (though this is changing, especially in the wake of the Israeli attack on the Osirak reactor in Iraq); and it has implemented an inspection system that often compromises strict national sovereignty. That safeguards are narrowly defined by IAEA as aiming to prevent diversion of fissile material, *not* to prevent establishment of latent capabilities for such diversion, is a weakness inherent in IAEA's structure. IAEA can do only what its members allow it to do. Until now IAEA's protocols define safeguards to cover actual diversion, not capability for diversion. Thus, for example, IAEA safeguards can deal with actual diversion of U²³⁵ in Pakistan from safeguarded reactors, but can do

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Bad news for IAEA — but the announcement of Israel's attack on Iraq's Osirak reactor on June 7, 1981 is greeted with celebrations on the streets of Tel Aviv.

complex matter. Most of Scheinman's study is therefore devoted to safeguards and the non-proliferation regime.

That regime, as described by Scheinman, includes much more than safeguards — that is, inspections and material accountability — *per se*. It includes a general world consensus against the spread of nuclear weapons; international undertakings for nuclear co-operation and trade; voluntary constraints by nuclear suppliers; the treaty of Tlatelolco and the NPT; verification of peaceful use and no-weapons pledges; and the IAEA itself to verify compliance with safeguards and to facilitate access to peaceful uses of nuclear energy. Safeguards are but one element of this regime, and cannot be expected to carry the full burden for preventing proliferation.

nothing about the Pakistan centrifuge plant.

In this somewhat narrow sense, Scheinman is correct in his generally positive assessment of IAEA. On the other hand, he hints that the safeguards inspection may have been less than adequate in the case of the Osirak; and that IAEA's ability to prevent non-NPT signatories, such as Pakistan, from acquiring nuclear explosives, is very limited.

Granted that IAEA is a remarkable international agency, the underlying question remains — is the non-proliferation regime as described in this book sufficient to break the connection between nuclear power and nuclear proliferation? As befits a scholar, Scheinman does not quite answer this central question. Nor is the question really

answerable. As judged by the small number of overt nuclear weapons states (only five plus India), the non-proliferation regime has worked far better than the Cassandras of the 1950s had predicted. But judging from the existence of 'threshold' states — Pakistan, Israel, South Africa — we may be premature in this assessment. Additional measures, beyond the existing regime, ought to be considered: Scheinman offers us several, without advocating any of them very strongly.

Such measures might include proscription of reprocessing except under IAEA supervision; sequestering of plutonium, either in centralized storage, or in unprocessed fuel in supervised waste repositories; and a halt to vertical proliferation, culminating in a comprehensive test ban (CTB). The first of these measures may come into being in many nuclear power systems simply because reprocessing is too expensive; the second, which in a way captures some of the flavour of the original Eisenhower Atoms-for-Peace proposal, continues to receive sporadic support. The CTB, which some regard as the talisman that will end the spectre of

proliferation (and which I strongly support), may be closer at hand, in the wake of the INF Treaty, than was conceivable at the time Scheinman wrote his book.

Can we, then, have nuclear power without proliferation, as nuclear proponents aver? Or must the world eventually forego nuclear power because proliferation is inevitably the unwanted bastard of fission? Professor Scheinman takes the optimistic view — that nuclear proliferation need not be the inevitable consequence of nuclear power provided the non-proliferation regime, and the IAEA, continue, little-by-little, to be strengthened. Whether additional states will 'go nuclear' by routes that do not require power plants and that circumvent the jurisdiction of IAEA, or of any international body for that matter, remains, and will always remain, an open, political question. □

Alvin M. Weinberg, Institute for Energy Analysis, PO Box 117, Oak Ridge, Tennessee 37830, USA, is a former Director of Oak Ridge National Laboratory. With John Barkenbus, he is co-editor of The Nuclear Connection: A Reassessment of Nuclear Power and Nuclear Proliferation (Paragon House Press, 1987).

Why do they do it?

Michael Bulmer

The Evolution of Sex and its Consequences. Edited by S.C. Stearns. *Birkhäuser*: 1987. Pp. 403. SwFr. 80, \$59.50.

The Evolution of Sex: An Examination of Current Ideas. Edited by Richard E. Michod and Bruce R. Levin. *Sinauer/Blackwell Scientific*: 1988. Pp. 342. Hbk \$55; pbk \$29.95. In the UK pbk £23, hbk price yet to be set.

SEX is the big problem in evolutionary biology, the one we should all like to solve. Sexual reproduction has two clear disadvantages. First, recombination, its main consequence, breaks up coadapted gene complexes, which must be a bad thing in a constant environment. Second, there is the two-fold cost of sex. In anisogamous species with no paternal care of the young, a female who produces only female offspring parthenogenetically has a two-fold advantage over a sexual female who produces equal numbers of male and female offspring. There is some verbal disagreement whether this should be attributed to the cost of producing useless males or to the cost of genome dilution by meiosis and syngamy, but its reality is beyond doubt.

Why then do so many plants and animals do it? The fact that recombination breaks up gene complexes suggests that it can only be advantageous in a variable environment, but how can the advantage

be large enough to outweigh the two-fold cost? This intriguing question has challenged theorists to produce an explanation and experimentalists to measure the relevant costs and benefits.

These two multi-author reviews are complementary, with remarkably little overlap either in authorship or coverage. The book edited by Stearns is rather broad in its scope, with chapters on the evolution of mating types and anisogamy, the costs of sex, sex-determining mechanisms, sex-ratio and sex-allocation theory, and sexual selection in animals and plants, organized around a central core which discusses the leading hypotheses for the evolutionary maintenance of sex and the evidence for and against them.

I enjoyed most of these contributions, particularly those dealing with the empirical evidence. Bierzychudek reviews experimental tests of the advantage of sexual over asexual progeny, mostly the work of Antonovics and his students on the grass *Anthoxanthum* in which a substantial advantage has been found for sexual progeny. A criticism of this work is that the asexual progeny are obtained from vegetatively produced tillers which are subject to parental contamination; to avoid this problem Bierzychudek is currently working with *Antennaria* which produces both sexual and asexual (apomictic) seed.

Hebert reviews the comparative evidence about the genetics of cyclical parthenogens, such as aphids and *Daphnia*, which throws light both on the consequences of going asexual for a few genera

tions and on the evolution of cyclical parthenogenesis. Cyclical parthenogens are often quoted as showing that there must be a short-term advantage for sex, because they could easily become obligate parthenogens. Hebert disputes this argument and concludes that "asexuality originates as a product of internal genetic factors rather than as a response to selection pressures imposed by the environment"; I doubt however whether this conclusion would stand up to theoretical analysis. Gouyon and Couvet review data on the fecundity of females in gynodioecious plants, such as thyme, in which females and hermaphrodites coexist, and discuss the evolutionary maintenance of this condition.

The book edited by Michod and Levin concentrates in greater depth on the central problem of the evolutionary maintenance of sex. It gains breadth by examining a diversity of hypotheses, and by its extensive consideration of the molecular basis of recombination. Two lucid chapters by Felsenstein and Maynard Smith review the orthodox views on the evolution of recombination, while Seger and Hamilton discuss the front-running 'Red Queen' model in which sex is regarded as an adaptation in the coevolutionary battle between hosts and their parasites.

In contrast, Bernstein *et al.* regard recombination as an adaptation for DNA repair rather than for generating variability, and Holliday presents a variant of this idea based on the maintenance of the correct pattern of DNA methylation in the germ line; I found these chapters stimulating but unconvincing. Hickey and Rose present the even more way-out idea that the function of sex is to facilitate the horizontal transfer of parasitic DNA. There is a thought-provoking chapter by Levin on the evolution of sex in bacteria.

In summary, Stearns's book is strong on the empirical evidence, Michod and Levin's on the competing theories for the evolutionary maintenance of sex. Where do they leave us? Felsenstein is cynical:

This year, the sex crisis seems to have returned . . . What has happened? Has a new source of data or a new kind of experiment been discovered that will help us resolve the controversies? . . . No. . . . The problem has simply flared up again and will probably gutter out after a while. Biologists will once again all become convinced that they know the answer, but once again there will be no unanimity as to what the answer turned out to be.

I disagree with this assessment. These two books, taken together, show that there is now much better understanding of the theoretical possibilities, and that experiments to distinguish between them are being done. □

Michael Bulmer is a Lecturer in the Department of Biomathematics, University of Oxford, Oxford OX1 3UB, UK.

Unveiling the face of Selene

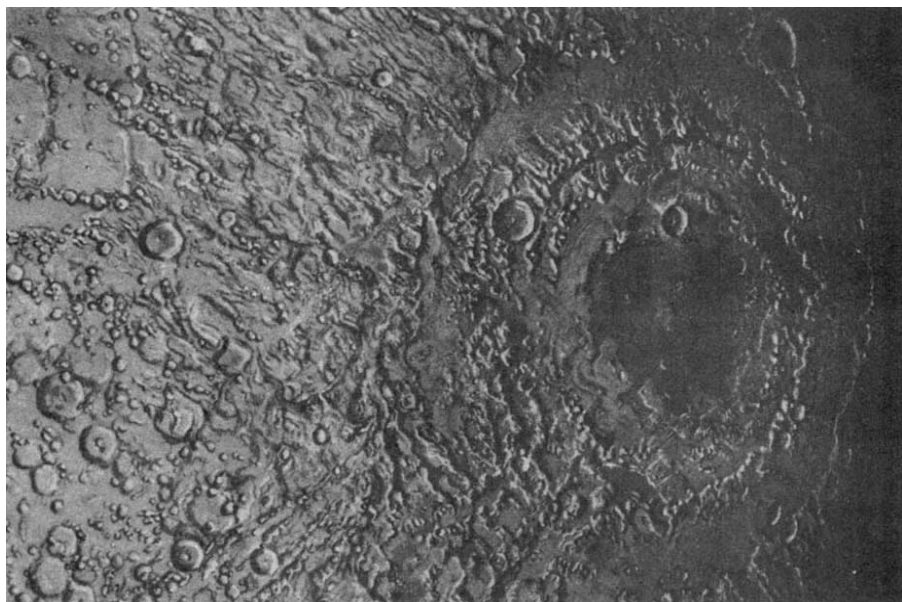
S.R. Taylor

The Geologic History of the Moon. US Geological Survey Professional Paper 1348. By Don E. Wilhelms with sections by John F. McCauley and Newell J. Trask. US Government Printing Office: 1987. Pp. 302 + 25 colour plates. \$33.

DON Wilhelms has put together the definitive study of the stratigraphy and geological history of the Moon. The chief glory of the book is in the carefully chosen and well-reproduced pictures, which are complemented by Wilhelms's sage commentary. Perceptive sections on Orientale by John McCauley and on crater morphology by Newell Trask add to the book's value. The text draws deeply on the accumulated wisdom of the USGS Astrogeology Branch in Flagstaff, Arizona. Wilhelms pays generous tributes both to his co-workers and to the pioneering work of Gene Shoemaker, based on his definitive study of Meteor Crater, Arizona, itself the subject of long debate among geologists obsessed with internal, rather than external processes.

Among the various insights to be gained from the book are that it is hazardous to take the Earth as an analogy for the Moon, confirming the view that much terrestrial geochemistry and petrology is unique to this planet; that simple rather than complex explanations of lunar features are usually correct; and that lunar craters are all effectively the result of impact (even the curious case of crater Kopff), despite the desperate attempts of proponents of explosive volcanism to turn them into volcanic calderas. It is salutary to read of all these past controversies and of the problems of interpretation of photographs in the absence of 'ground truth'.

Some controversies remain unresolved. The nature of the Apennine Bench Formation, whether impact-generated, analogous to the Montes Rook Formation of the Orientale Basin, or composed of KREEP basalt, beloved of petrologists, must await further missions. The lunar cataclysm, a long-lived but erroneous view of lunar cratering history, is dismissed in one succinct paragraph (pp. 190–191). Wilhelms also comments mildly on the petrological versus the stratigraphical approach; both disciplines need each other. Thus "names derived from terrestrial igneous petrology have been applied, inappropriately, to rocks melted and emplaced by impacts" (p. 139). In this context, Wilhelms's careful delineation of the distinction between impact basins and the later voluminous lavas which fill them



The young one — Orientale basin and its deposits, from a painting by Donald E. Davis.

to form the maria should help to remove another lingering misconception.

Readers will be enlightened on many topics: the problems of using lunar craters to date surfaces (common sense abounds in the treatment of this difficult technique); the origin of the sinuous rilles; the misconceptions about lunar structure and tectonics which led to the 'lunar grid', and the final realization that the crustal structure is dominated by the effects of the giant ringed basins, which are themselves the result of asteroidal or planetesimal impact.

It is worth contemplating these structures (of which the type example is the youngest, the 900-km diameter Orientale Basin dating from about 3.8×10^9 yr ago). The formation of multi-ring basins is ascribed to fluidization, rather than slumping into a deep initial cavity. Their effects outside the basin itself are remarkable: "a continuous deposit extends outward for an average distance of 450 km or one basin radius from Montes Cordillera [the outer ring of the Orientale basin]. . . . In some sections, dense clusters of secondaries (craters) extend 900 km (two basin radii) from the Cordillera. Secondaries also extend beyond 1850 km in at least one ray-like string" (p. 66). This enormous structure, with all its complex stratigraphy, formed in minutes.

Evidence of such processes must give pause to those seeking rocks older than 3.8×10^9 yr on the Earth. Current estimates indicate that at least 200 such basins, with diameters greater than 1,000 km, formed on Earth in the interval $4.4\text{--}3.8 \times 10^9$ yr ago, thus explaining the absence of identifiable terrestrial rock units. The impact melts and breccias from this epoch would have been easy meat for terrestrial eroding agents; in contrast, the lunar examples are still preserved on the bone-dry Moon.

Perhaps the most notable aspect of the book is the emphasis on the importance of the Procellarum Basin (originally perceived as the Gargantuan Basin by Peter Cadogan), a stupendous structure 3,200 km in diameter, within which nestle most of the later near-side basins and mare basalt flows. This basin may be responsible for the thinness of the near-side lunar crust, the prevalence of near-side mare basalt, the widespread flooding of Oceanus Procellarum and much more.

Wilhelms wisely refrains from discussing the origin of the Moon, a question which has been resolved only since the completion of the book. But his endeavours to place lunar stratigraphy on a sound footing have helped to place constraints on the problem.

Full justice to the photographs and drawings is done by the generous format (28×37 cm). But this is a coffee-table book with a difference, for there are ten pages of references which constitute a very useful lunar bibliography. The cover is enhanced by a sequence of drawings, well worth close study, by Don Davis, who also provides other illustrations through the book. A final delight at the end of the text is the sequence of 24 superb colour plates, summarizing ringed basins, maria, and lunar structure and stratigraphy.

There is so much accumulated wisdom in this book that a reviewer can only hope to provoke readers to delve into it for themselves. It should serve as a model for similar studies of the other planets and satellites. As Wilhelms comments (p. 139), "the lunar pre-Nectarian System provides our closest look at the early Solar System", and that is probably our ultimate scientific justification for going to the Moon. □

S.R. Taylor is at the Research School of Earth Sciences, Australian National University, GPO Box 4, Canberra 2601, Australia.

Those old familiar edifices

Fiona Steele

The Great Engineers. Edited by Derek Walker. *Academy Editions, 42 Leinster Gardens, London/St Martin's Press, New York: 1987. Pp.288. £35, \$50.*

It was fitting, and indeed encouraging, that the 150th anniversary of the founding of the Royal College of Art (RCA) should have been marked recently by a compelling exhibition devoted to a celebration of engineers and engineering, and by the publication of this book. The thesis embodied in *The Great Engineers*, a volume of some 20 essays edited and co-authored by Professor Derek Walker, is the perpetual search for excellence which has dominated the lives of Britain's greatest inventors and engineers, and the relationship between art and design, and manufacture and construction.

The familiar names from the past are here — Brunel, Telford, the Stephenson — but Walker has encouraged his contributors to provide very personal perspectives of the lifestyles and achievements of their chosen subjects. Although these engineers attracted wide public attention and renown, there are others, perhaps not so well known, whose contributions were equally significant. So, for example, we have an essay on Sir Joseph Bazalgette, who did so much to improve the health of Londoners through his design and execution of a massive programme of sewer installation in the last century. This is one of the most interesting contributions in the book and is especially timely in view of the present debate over infrastructure renewal, a problem not confined to Britain. It is salutary to learn that throughout his career Bazalgette argued for powers to bring private water companies in London under the control of a single Metropolitan Water Board; and not content with the problems of water supply and sewage disposal, he also sought to improve the total built environment and laid down criteria for planning which are still pertinent today.

Moving into the twentieth century, the legend of excellence is represented by the great modern partnerships, Arup, Samuely, Happold, where engineering and art continue to feed on each other to produce visually exciting, if sometimes controversial, structures. This marriage of minds brought together through design is a recurring message throughout the various contributions.

Not that the book is purely concerned with the 'structural' side of things. Walker has acknowledged the diversity of engineering by commissioning contributions on

communications, materials and electronics. Although these provide an interesting balance, they do not offer the excitement and infectious enthusiasm of the rest of the book. The RCA itself is featured in an intriguing, indulgent article on the Duke of Wellington's funeral car, and indeed the whole book is very much an indulgence. But celebrations are a legitimate excuse for some excess, and the book deserves a wider reading and viewing than the engineering fraternity. It is beautifully illustrated, looking at structures in their entirety and in the context of their setting, and then delving into the minutiae of their

anatomy. It must have been great fun assembling the pictures and drawings.

There are niggles — duplication in illustrations, a lack of order, too much text in one place, too little in another, non-British inventions in a list of British inventions — but these can be tolerated. The book is not easily portable but if taken on a journey readers must be prepared to accommodate the over-the-shoulder attention of their fellow passengers. □

Fiona Steele is Head of the Technology Group, Confederation of British Industry, Centre Point, 103 New Oxford Street, London WC1A 1DU, UK.

Circular reasoning

Thomas J. Kindt

Regulatory Idiotopes. By Constantin A. Bona. *Wiley: 1987. Pp.279. \$59.95, £52.50.*

It is elementary that injection of an animal with an antigen results in formation of an antibody (AB1). When a genetically matched animal is then injected with AB1, an anti-antibody (AB2) directed against unique determinants of AB1 is generated. AB2 is generally termed an anti-idiotype.

What happens when AB2 is injected into a third genetically identical animal ceases to be elementary. If, for example, the antigen that induced AB1 was a hormone, then AB2 may possibly react with the receptor for that hormone because AB2 contains the 'internal image' of the original antigen. Continuing along the same line of reasoning, AB2 can be used as a vaccine to generate a response to a harmful pathogen and thereby circumvent the need to expose the subject in order to achieve immunity. Similarly, memory may be carried in the immune system by antibodies that have internal images of previously encountered antigens. All this may sound like an immunologist's fantasy, but most of these manipulations have been carried to successful conclusions and constitute the subject of Constantin Bona's book.

Although the concept of unique determinants on immunoglobins has been in the literature since 1955, and the word 'idiotype' has been in use since 1963, only a few early investigators experimented with this tool. It took the ideas embodied in Niels Jerne's network hypothesis of the early 1970s to elicit broad interest in the possibilities that idiotypy had to offer. The 'idiotope' in the title of this book is the epitope of an idiotype, and the fact that it is modified by 'regulatory' implies that there are other (non-regulatory) idiotopes. In spite of this qualification, Bona satisfactorily develops the general concepts necessary to view idiotypy in proper

perspective. A certain sophistication is assumed on the part of the reader and the introduction would have been better for the distinction used by Oudin to separate antigenic determinants of antibodies into isotype, allotype and idiotype. This information has escaped many modern students of immunology, who tend to view antibodies (B-cell receptors) as products of interacting genes rather than as serological targets.

This monograph takes the reader beyond the basic concept of idiotypy, concentrating throughout on the premise that the idiotope is a central player in the biological drama, a regulatory element with a myriad of possible functions. Although the treatment is not without some bias towards Bona's own considerable contributions to the area, his insights are keen and will help guide the reader confronted with a distillation of over 600 references. It is not possible to digest the entire volume in one sitting, but the chapters and subchapters stand alone quite well and some are entertaining as well as informative. That covering internal images was especially appealing to me. Other chapters are too crammed with examples to make light reading, but will serve as excellent reference material on the subject matter concerned.

The book is the second volume of a series edited by Bona, and will be a worthwhile acquisition for those who have a direct or peripheral interest in the area. Even those who disagree with the concept of regulation by antibodies will find it to be an excellent source of information. For myself, *Regulatory Idiotopes* will replace an overflowing file of papers on the subject, and will serve as a guide to a field that I have monitored with varying degrees of enthusiasm for many years. □

Thomas J. Kindt is Chief of the Laboratory of Immunogenetics, National Institutes of Health, Bethesda, Maryland 20892, USA.

• Erratum. The recently published second edition of *The Idea of Prehistory* is by Glyn Daniel with two additional chapters by Colin Renfrew. The first edition was by Glyn Daniel, not by Colin Renfrew as suggested (*Nature* 331, 493; 1988).

Peptide growth factors are multifunctional

Michael B. Sporn & Anita B. Roberts

Laboratory of Chemoprevention, National Cancer Institute, Bethesda, Maryland 20892, USA

The actions of many peptide growth factors include both stimulation and inhibition of cell proliferation, as well as effects unrelated to the control of cell growth. One peptide can have both stimulatory and inhibitory activity in a single cell, depending on the context of the other signal molecules present.

MANY peptides with potent stimulatory effects on proliferation of either epithelial or mesenchymal cells have been identified in the past ten years; recently peptides with strong inhibitory effects on cell proliferation have also been found. Because of their regulatory action on tissue growth these peptides have been termed growth factors, most having been named according to the biological activity and the assays leading to their original isolation, examples being epidermal growth factor, fibroblast growth factor and transforming growth factors. It is now becoming evident that many of these peptides have a much wider range of action than the biological activity for which they were originally named, and most of them are better considered as multifunctional agents. Table 1 provides examples of peptides that are commonly known as growth factors and have three separate actions: stimulation of cell proliferation, inhibition of cell proliferation, and effects on cell function apparently unrelated to proliferation.

Range of cellular action

Not only may the actions of an individual peptide growth factor on a given cell type be diverse, but the range of cells which respond to the peptide may also be much wider than originally determined. The set of peptides known as interleukins are a good example. They were first defined as signalling molecules controlling activities of cells in the immune system. Thus, interleukin-2 (IL-2) was originally named T-cell growth factor; but it is now known that it is also an important growth factor for B cells^{20,21} and increases the cytotoxic activity of monocytes²⁴. Conversely, interleukin-4 (IL-4) was first characterized as a B-cell growth factor (BSF-1) (ref. 32), but recently it has been

shown to be a potent stimulant of T-lymphocytes^{33,34} and to control the maturation of other haematopoietic cells, such as granulocytes and macrophages³⁵. Indeed, the concept of an interleukin is somewhat nebulous now that it has been shown that IL-1, first described as a substance produced by macrophages that acts on lymphocytes, is also a potent stimulator of epidermal keratinocytes, chondrocytes and fibroblasts¹⁶, and can inhibit the growth of breast cancer cells¹⁷; also, IL-2 may control the growth of glial cells of the nervous system^{22,23} as well as the proliferation of T and B lymphocytes^{20,21}. Conversely, peptides originally isolated from non-immune cells and defined by their actions on non-immune cells have now been found to have potent effects on the immune system. Examples are transforming growth factor-beta (TGF- β) and neuroleukin. Thus, TGF- β is at least 10^4 times more potent (on a molar basis) than the peptide cyclosporin A as a suppressant of T lymphocytes¹⁵. Neuroleukin, a peptide isolated from denervated muscle and salivary glands and neurotrophic for spinal and sensory neurons³⁶, can induce maturation of B-lymphocytes and stimulate their secretion of immunoglobulin at picomolar concentrations³⁷. Both TGF- β and neuroleukin are also synthesized by T lymphocytes after stimulation with lectins^{13,37}. Another example is interleukin-6 (IL-6, also known as interferon- β 2 and B-cell stimulatory factor 2, BSF-2), which is synthesized in non-immune cells such as fibroblasts and various human tumour cells and has profound paracrine activities on B lymphocytes, increasing immunoglobulin synthesis and secretion^{27,28,38}. TGF- β , neuroleukin, and IL-6 thus may all provide bidirectional communication between cells of the immune system and those outside it²⁷.

Table 1 Peptide growth factors are multifunctional

Growth factor	Proliferative effects	Anti-proliferative effects	Effects unrelated to proliferation
Epidermal growth factor	Keratinocytes, fibroblasts ¹	Hair follicle cells ² , squamous carcinoma cells ³	Suppression of gastric acid secretion ⁴
Fibroblast growth factor (basic)	Many mesenchymal cells, especially endothelial cells ⁵⁻⁷	Ewing's sarcoma cells, other tumour cells ⁸	Regulation of pituitary and ovarian cell function ⁵
Transforming growth factor-beta	Fibroblasts ^{9,10} , osteoblasts ^{11,12}	Fibroblasts ⁹ , epithelial cells ¹⁰ , T-lymphocytes ¹³ , osteoblasts ¹²	Suppression of immunoglobulin secretion ¹⁴ and adrenal steroidogenesis ¹⁵
Interleukin-1	Lymphocytes, keratinocytes, fibroblasts ¹⁶	Breast cancer cells ¹⁷	Endogenous pyrogen ¹⁸ , bone resorption ¹⁹
Interleukin-2	T- and B-lymphocytes ^{20,21} , oligodendrocytes ²²	T-lymphocytes ²¹ , oligodendrocytes ²³	Increased cytotoxic activity of monocytes ²⁴
Interleukin-6	Hybridoma, plasmacytoma cell lines ²⁵	Breast cancer cells ²⁶	Increased expression of cell surface antigens ²⁷ and secretion of immunoglobulins ²⁸
Tumour necrosis factor-alpha	Diploid fibroblasts ^{29,30} , epithelial cells ²⁹	Many tumour cells ²⁹	Inhibition of lipoprotein lipase, stimulation of collagenase ³¹

The table is not intended to be comprehensive; examples have been chosen to illustrate the text.

Another set of effector molecules acting on a wide variety of cell types are the regulatory neuropeptides, such as vasopressin, bombesin, substance P, substance K and β -endorphin. Although primary interest in these peptides has centred on their roles in the nervous system, both central and peripheral, they also have significant mitogenic activity on cells as diverse as chondrocytes, bronchial epithelial cells, smooth muscle cells, fibroblasts and lymphocytes, thus enabling the nervous system to exert control of cell proliferation in a variety of distant target sites^{39,40}.

Interaction between growth factors

Even in a single cell type, the nature of growth-factor action may depend on the context set by other substances present: for example, TGF- β stimulates growth of certain fibroblasts *in vitro* in the presence of platelet-derived growth factor (PDGF) but inhibits their growth if epidermal growth factor (EGF) is present⁴¹. TGF- β can also stimulate replication of osteoblasts^{11,12} but its mitogenic action can be reversed by other peptide growth factors such as EGF, PDGF, and tumour necrosis factor- α (TNF- α), which themselves promote DNA synthesis in osteoblasts⁴². The action of a growth factor can also depend on the state of differentiation of a target cell. For example, TGF- β stimulates or inhibits the expression of the cartilaginous phenotype in embryonic mesenchyme or chondroblasts, according to the developmental stage of the cells: early in development, type II collagen is induced⁴³, but later it is suppressed⁴⁴.

The importance of context in determining the action of peptide growth factors is just beginning to be appreciated, but has recently also been demonstrated for IL-4 and TNF- α . IL-4 can either enhance or antagonize the growth of an entire group of haematopoietic progenitor cells, depending on the other growth factors present, notably interleukin-3 (ref. 35). The action of TNF- α is often closely coupled to two other peptide mediators, IL-1 and interferon- γ , in several complex ways^{31,45}. Furthermore, TGF- β and TNF- α each reverse the effects of the other on T-lymphocytes⁴⁶. Moreover, in BALB-3T3 cells TNF- α can be primarily mitogenic, primarily cytotoxic, or both mitogenic and cytotoxic in the same cell line, depending on the culture conditions, the concentration of TNF- α , and the presence or absence of other growth factors, particularly EGF and PDGF (J. Vilček and V. Palombella, personal communication). Similar interactions might be expected to operate between other peptide growth factors.

Interaction between receptors

The cellular mechanisms responsible for the multiple actions of a peptide ligand remain essentially unknown and there is no adequate explanation for the observed context-dependent effects. One mechanism that is probably important is the ability of a receptor for a specific peptide to alter either the cellular distribution or the binding affinity of the receptor for a second peptide growth factor, independent of any direct cross-reactivity of the peptides themselves. For example, insulin does not bind to the receptor for insulin-like growth factor-II (IGF-II), but does cause the cycling and redistribution of IGF-II receptors^{47,48}. Also PDGF can decrease the affinity of the EGF receptor for its ligand^{49,50}. This phenomenon is termed transmodulation (receptor-cross-talk) to distinguish it from the reduction in receptor numbers caused by homologous ligands, known as down regulation or down modulation.

Receptor molecules themselves can also be multifunctional, as shown recently for the receptor for IGF-II: this appears to be structurally and biochemically identical to the mannose-6-phosphate receptor^{51,52}. Thus separate binding sites for two distinct ligands have been found in a single receptor molecule. It has been suggested that the ligands for the mannose-6-phosphate binding sites are lysosomal enzymes, which somehow direct the intracellular fate of IGF-II when the ligand-receptor complex is internalized⁵¹.

A signalling language

What sense can be made of the multiplicity of actions of so many peptide growth factors? It is apparent that specific peptides are not necessarily limited to a single physiological activity, but rather that growth factors form part of a complex cellular signalling language, in which the individual peptides are the equivalent of characters of an alphabet or code. From this it follows that informational content resides not in individual peptides, but in the pattern or set of regulatory peptide molecules to which a cell is exposed. The use of combinations selected from a large number of peptide signalling molecules increases the amount of information that can be transmitted. It is not surprising that meaning in this particular cellular language is contextual, because that is generically the case for all languages or codes. Furthermore, the redundancy of individual peptides, or sets of peptides, as effective mitogenic signals in a given cell may promote error-free transmission of information in the presence of noise⁵³⁻⁵⁴; an example is seen in the numerous combinations of peptides and related signals that can elicit DNA synthesis in 3T3 cells⁵⁵.

The mechanisms for the integration of the great variety of signals that are transmitted from specific cell surface receptors to the interior of the cell remain perhaps the most difficult problem in this field. How does a cell read sense from the peptide growth factor alphabet? The solution to this question will be of major importance in cell biology in the coming decade. The observation that FGF is itself rapidly translocated to the nucleus, where it stimulates the transcription of ribosomal genes⁵⁶, raises the possibility that nuclear mechanisms for integration will be critical in determining the actions of peptide growth factors⁵⁷.

Conclusions

The original context of the discovery of particular signalling peptides has seldom indicated the extent of their diverse range of action. The highly specific fit of a peptide ligand with a glycoprotein cell membrane receptor provides a modular regulatory element that the evolutionary process has repeatedly used in different types of cells for many different purposes, including the alteration of their growth or differentiation. The action of any peptide signalling molecule will be defined ultimately in the context of its receptor and by the way in which signals from a set of receptors are integrated by the cell itself. Finally, the realization that the action of peptide growth factors depends on their context should lead to more effective therapeutic application, as their effects can be optimized in appropriate combinations. Recombinant technology will increase the availability of growth factors for such purposes.

We thank Tony Burgess, Igor Dawid, Victor Fried, Gordon Gill and Jan Vilček for helpful comments.

1. Carpenter, G. & Cohen, S. A. *Rev. Biochem.* **48**, 193-216 (1979).
2. Panaretto, B. A., Leish, Z., Moore, G. P. & Robertson, D. M. *J. Endocr.* **100**, 25-31 (1984).
3. Kamata, N. *et al. Cancer Res.* **46**, 1648-1653 (1986).
4. Gregory, H. *Nature* **257**, 325-327 (1975).
5. Baird, A. *et al. Recent Prog. Horm. Res.* **42**, 143-205 (1986).
6. Folkman, J. & Klagsburn, M. *Science* **235**, 442-447 (1987).
7. Gospodarowicz, D. *Nucl. Med. Biol.* **14**, 421-434 (1987).
8. Schweigerer, L., Neufeld, G. & Gospodarowicz, D. *J. clin. Invest.* **80**, 1516-1520 (1987).
9. Sporn, M. B., Roberts, A. B., Wakefield, L. M. & de Crombrughe, B. *J. Cell Biol.* **105**, 1039-1045 (1987).
10. Moses, H. L. *et al. Cancer Cells* **3**, 65-71 (1985).
11. Robey, P. G. *et al. J. Cell Biol.* **105**, 457-463 (1987).
12. Centrella, M., McCarthy, T. L. & Canalis, E. C. *J. biol. Chem.* **262**, 2869-2874 (1987).
13. Kehrl, J. H. *et al. J. exp. Med.* **163**, 1037-1050 (1986).
14. Kehrl, J. H. *et al. J. Immunol.* **137**, 3855-3860 (1986).
15. Hotta, M. & Baird, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7795-7799 (1986).
16. Durum, S. K., Schmidt, J. A. & Oppenheim, J. J. A. *Rev. Immunol.* **3**, 263-287 (1985).
17. Gaffney, E. V. & Tsai, S.-C. *Cancer Res.* **46**, 3834-3837 (1986).
18. Murphy, P. A., Simon, P. L. & Willoughby, W. F. *J. Immunol.* **124**, 2498-2501 (1980).
19. Gowen, M. & Mundy, G. R. *J. Immunol.* **136**, 2478-2482 (1986).
20. Waldmann, T. A. *et al. J. exp. Med.* **160**, 1450-1466 (1984).
21. Taniguchi, T. *et al. Immunol. Rev.* **92**, 122-133 (1986).
22. Benveniste, E. N. & Merrill, J. E. *Nature* **321**, 610-613 (1986).
23. Saneto, R. P., Altman, A., Knobler, R. L., Johnson, H. M. & de Vellis, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9221-9225 (1986).
24. Malkovsky, M. *et al. Nature* **325**, 262-265 (1987).
25. Van Damme, J. *et al. J. exp. Med.* **165**, 914-919 (1987).
26. Revel, M. *et al. J. Leukocyte Biol.* **42**, 554 (1987).

27. Sehgal, P. B., May, L. T., Tamm, I. & Vilček, J. *Science* **235**, 731–732 (1987).
28. Kishimoto, T. & Hirano, T. A. *Rev. Immun.* **6**, 485–512 (1988).
29. Sugarman, B. J. *et al. Science* **230**, 943–945 (1985).
30. Vilček, J. *et al. J. exp. Med.* **163**, 632–643 (1986).
31. Beutler, B. & Cerami, A. *Nature* **320**, 584–588 (1986).
32. Paul, W. E. & Ohara, J. A. *Rev. Immun.* **5**, 429–459 (1987).
33. Mosmann, T. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 5654–5658 (1986).
34. Fernandez-Botran, R., Krammer, P. H., Diamenstein, T., Uhr, J. W. & Vitetta, E. *J. exp. Med.* **164**, 580–593 (1986).
35. Rennick, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6889–6893 (1987).
36. Gurney, M. E., Heinrich, S. P., Lee, M. R. & Yin, H.-S. *Science* **234**, 566–574 (1986).
37. Gurney, M. E. *et al. Science* **234**, 574–581 (1986).
38. Poupart, P. *et al. EMBO J.* **6**, 1219–1224 (1987).
39. Hanley, M. R. *Nature* **315**, 14–15 (1985).
40. Zachary, I., Woll, P. J. & Rozengurt, E. *Dev. Biol.* **124**, 295–308 (1987).
41. Roberts, A. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 119–123 (1985).
42. Centrella, M., McCarthy, T. L. & Canalis, E. *FASEB J.* **1**, 312–317 (1987).
43. Seyedin, S. M. *et al. J. biol. Chem.* **261**, 5693–5695 (1986).
44. Rosen, F. M., Stempien, S. A., Thompson, A. Y. & Seyedin, S. M. *J. cell. Physiol.* (in the press).
45. Dinarello, C. A. *et al. J. exp. Med.* **163**, 1433–1450 (1986).
46. Ranges, G. E., Figari, I. S., Espevik, T. & Palladino, M. A. Jr. *J. exp. Med.* **166**, 991–998 (1987).
47. Corvera, S. & Czech, M. P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7314–7318 (1985).
48. Rechler, M. M. & Nissley, S. P. A. *Rev. Physiol.* **47**, 425–452 (1985).
49. Bowen-Pope, D. F., DiCorleto, P. E. & Ross, R. J. *Cell Biol.* **96**, 679–683 (1983).
50. Zachary, I. & Rozengurt, E. *Cancer Surv.* **4**, 729–765 (1985).
51. Morgan, D. O. *et al. Nature* **329**, 301–307 (1987).
52. Lobel, P., Dahms, N. M., Breitmeyer, J., Chirgwin, J. M. & Kornfeld, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2233–2237 (1987).
53. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).
54. Ashby, W. R. *An Introduction to Cybernetics* (Wiley, New York, 1956).
55. Rozengurt, E. *Science* **234**, 161–166 (1986).
56. Bouche, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6770–6774 (1987).
57. Burwen, S. J. & Jones, A. L. *Trends biochem. Sci.* **12**, 159–162 (1987).

ARTICLES

Non-seasonal changes in total column ozone from satellite observations, 1970–86

Donald F. Heath

Laboratory for Atmospheres, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Analyses of non-seasonal changes in total column ozone from satellite observations during the period 1970–86 suggest, when combined with analyses of Dobson network data, that the largest reduction in global ozone since 1959 occurred between 1978 and 1986. The existence of an Arctic region of enhanced ozone depletion with similar but less pronounced characteristics of the Antarctic region is described. Before 1983 large rates of decrease were confined to high latitudes, subsequently regions of large rates of ozone depletion have appeared at mid-latitudes.

CONCERN about anthropogenic threats to the ozone layer began with Johnston¹ in 1971 who suggested that nitrogen oxides deposited by supersonic transport aircraft in the stratosphere could reduce ozone to unacceptable levels. Subsequently Molina and Rowland² suggested in 1974 that free chlorine released from the photodissociation of manmade chlorofluoromethanes (CFMs) could also significantly increase the catalytic destruction of ozone. Farman *et al.*³ suggested in 1985 that the enhanced rates of ozone depletion observed in Antarctica during the second half of the 1970s were the result of the sensitivity of the Antarctic stratosphere to the growth of inorganic chlorine.

Currently two sources of data are available for the investigation of long-term changes in total column ozone on a global scale. These are the ground-based Dobson spectrophotometers and the satellite-borne instruments measuring ultraviolet solar and backscattered solar radiation. During the period of the International Geophysical year (1957–59) a global network of about one hundred Dobson spectrophotometers was established. They are located on land, and only about 20% of these ground-based instruments are located in the Southern Hemisphere. From 1970 measurements of total column and vertical distributions of stratospheric ozone have been obtained from polar-orbiting satellites.

Total ozone is determined from direct sun measurements of relative intensities of selected double pair wavelengths in the region from 305.5 to 339.8 nm from Dobson spectrophotometers⁴. The Nimbus 4 version of the solar backscattered ultraviolet (SBUV) instrument, the BUV, which operated from 1970 to 1977 and the Nimbus 7 SBUV instrument which has operated continuously since 1978 measure solar radiation backscattered from the atmosphere and solar irradiances at the top of the atmosphere at 12 discrete wavelengths from 255–340 nm^{5,6}. Computed atmospheric reflectivities between 255 and 306 nm are used in the profile inversion⁷ and between 312 and 340 nm are used to derive total ozone⁸. Both Nimbus satellites are sun-synchronous polar orbiters that cross the Equator near

local noon and provide ozone data from 81° S to 81° N during daylight, up to solar zenith angles of 88°. Recently both ozone data sets have been reprocessed with a common algorithm and are available from the National Space Science Data Center.

Here we report the analyses of long-term changes in total ozone derived from satellite observations during the period 1970–86 with the greatest emphasis placed on the analyses of observations obtained from the SBUV instrument on the Nimbus 7 satellite during the period November 1978 to October 1986. The Nimbus 7 spacecraft also carries the total ozone mapping spectrometer (TOMS) instrument which shares certain electronic, mechanical and optical components with the SBUV instrument⁶. Results from the TOMS instrument have not been included here because the TOMS ozone data had not been calibrated for the determination of trends at the time when these trend analyses were done. Here I do not attribute the long-term changes in total ozone from satellite observations to specific physical mechanisms. Rather, a description is given of the results of analyses of the changes in total ozone and their consistency with similar analyses of total ozone obtained from the Dobson network. The results of the analyses of ozone changes derived from satellite observations are compared with current model predictions of ozone changes resulting from increasing levels of anthropogenic constituents which catalytically deplete ozone. While a drift in instrument sensitivity with time in orbit is a major problem for satellite observations, it is suggested that there may be reasons other than an incorrect assessment of changes in instrument sensitivity which may contribute to the observed drift of the total ozone measurements from the SBUV instrument relative to measurements from the Dobson ground-based network.

Analyses of the temporal variability of the global budget and spatial distribution of total ozone during the period 1970–86 suggest possible significant influences of natural and anthropogenic perturbations through photochemical, dynamical and radiative processes extending from the polar regions to the

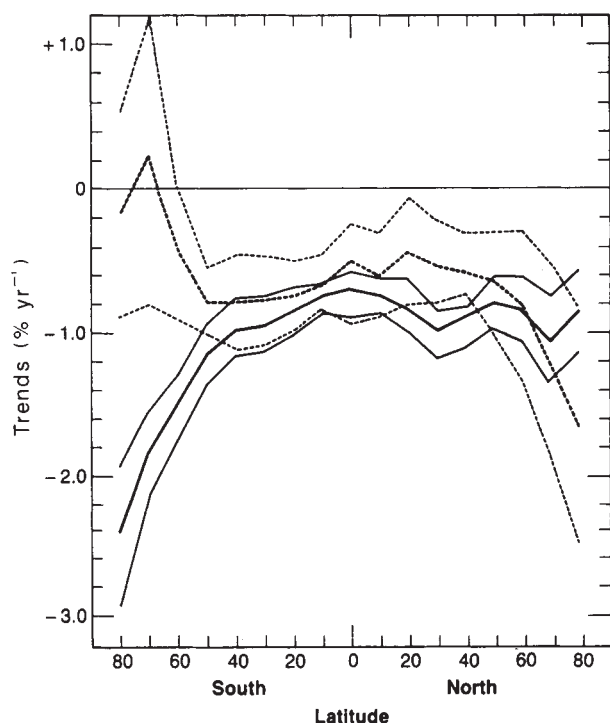


Fig. 1 Linear trends of total column ozone from the Nimbus 4 BUV instrument (heavy broken line) with 95% confidence limits (light broken lines) and the Nimbus 7 instrument (heavy solid line) and 95% confidence limits (light solid lines). No correction applied for possible instrument drift relative to Dobson network.

tropics. While a report⁹ on the status of ozone and temperature trends gives no significant overall total ozone trend in data from the Dobson network during this period, it was noted, however, that the inclusion of 1983 data made the trends slightly more negative.

During the 16-year period of satellite observations of total ozone from 1970 to 1986 the observed non-seasonal changes in total ozone can be influenced by errors in the assessment of drift in instrument sensitivity, the effects of variations in solar output during solar sunspot cycles 20 and 21, the eruptions of the volcano El Chichon in April 1982, dynamical changes associated with the El Niño–Southern Oscillation events of 1982–83 and 1972–73 and other climatic variations in circulation and temperature structure.

The principal method for assessing the effects of instrument drift and calibration errors is through comparisons of ozone determined from the satellite instruments and from the Dobson network. The evaluation of satellite instrument calibration drift in orbit by comparing with the ozone data obtained from the Dobson network may be in error due to a variety of reasons which are discussed in a subsequent section.

The observed non-seasonal changes in total ozone during the period from 1970–86 are described in terms of simple linear regression analyses against time, of deviations of monthly mean ozone from monthly means averaged over extended periods. Good global coverage with the Nimbus 4 instrument was obtained only during the first two years after launch. However, this coverage was not uniform at high latitudes in the Northern Hemisphere due to an early failure of a tape recorder which produced gaps associated with the playback of recorded data in the vicinity of the Alaska ground station. In subsequent years up to the termination of observations in early 1977, spatial coverage deteriorated due to spacecraft power problems for the instruments. Thus the analyses of Nimbus 4 ozone data were restricted to linear trends (1970–76), a comparison with the linear trends derived with the SBUV observations, 1978–86, and

a comparison of the seasonal and latitudinal variation of the trends with those from Nimbus 7. The operating characteristics of the Nimbus 7 instrument are much improved over the Nimbus 4 instrument in terms of stability, signal-to-noise, uniformity of spatial coverage and length of record. Superior operating characteristics of the Nimbus 7 SBUV instrument and reports of large ozone changes in the Antarctic after the mid-1970s are reasons for the much more extensive analyses of these more recent data.

Analyses covering an eight-year period from November 1978 to October 1986 have been made of monthly deviations from mean values averaged over six years (November 1978–October 1986 in five climate regions: polar, 55–81°; temperate, 25–55°; and tropical, 0–25°). The ozone data in the five climate regions are produced by an area-weighting of the monthly zonal mean data in 10° latitude bands centred at 10° intervals from 80° S to 80° N. The six-year monthly mean ozone data that are nearly centred about the ultraviolet maximum of solar cycle 21 are used to deseasonalize the monthly mean data used in the analyses.

The results of the analyses are used to show the existence of patterns of ozone changes as related to month, climate region and hemisphere. Those patterns are consistent with recent two-dimensional model calculations of photochemically induced reductions in total ozone²⁴. Similar analyses of deviations of area-weighted 10° monthly zonal mean data averaged from 65° S to 65° N indicate that global ozone (65° S–65° N) has decreased significantly over the eight-year period 1978–86. The decreases inferred from satellite observations in total ozone in the five climate regions and over the globe are similar to those derived from the analyses of Dobson network data. Derived geographical patterns of ozone changes in deseasonalized ozone data indicate the existence of regions of increasing rates of ozone depletion with increasing latitude in both hemispheres which are more longitudinally asymmetric in the Northern Hemisphere than in the Southern Hemisphere.

Observations

Latitudinal trends. Latitudinal dependence of trends in total ozone are computed from linear regression analyses of monthly deviations from seasonal means derived from observations with the Nimbus 4 BUV instrument (1970–76) and the Nimbus 7 SBUV instrument (1978–86). Seasonal means for the Nimbus 4 observations are computed from monthly mean values of total ozone for 10° latitude bands centred at 10° latitude intervals from 80° S to 80° N for the period May 1970–April 1982. A seasonal mean is derived from observations over only a two-year period because of missing data resulting from degradation of the power system and the failure of a tape recorder on the Nimbus 4 satellite. Corresponding monthly mean values of total ozone from Nimbus 7 observations were computed for an interval of six years from November 1978 to October 1984. The mid-point of this six-year total ozone data used to deseasonalize monthly mean values of total ozone nearly coincides with the maximum in UV solar spectral irradiance of solar cycle 21 during the period August–September 1981.

These trends are shown in Fig. 1 for Nimbus 4 observations (1970–76) by the heavy dashed line and for the Nimbus 7 observations (1978–86) by the heavy solid lines. The 95% confidence limits for the linear trends are given by corresponding dashed and solid thin lines. The linear trends shown in Fig. 1 have not been corrected for an apparent drift of the BUV

Table 1 Drift of BUV, SBUV instruments with respect to Dobson network and 95% confidence limits of the uncertainty in drift

	BUV–Dobson		SBUV–Dobson	
Period	1970–76	1978–84	1978–85	1978–86
No of stations	48	41	41	41
Drift (% yr ⁻¹)	-0.43 ± 0.16	-0.38 ± 0.13	-0.41 ± 0.11	-0.44 ± 0.12

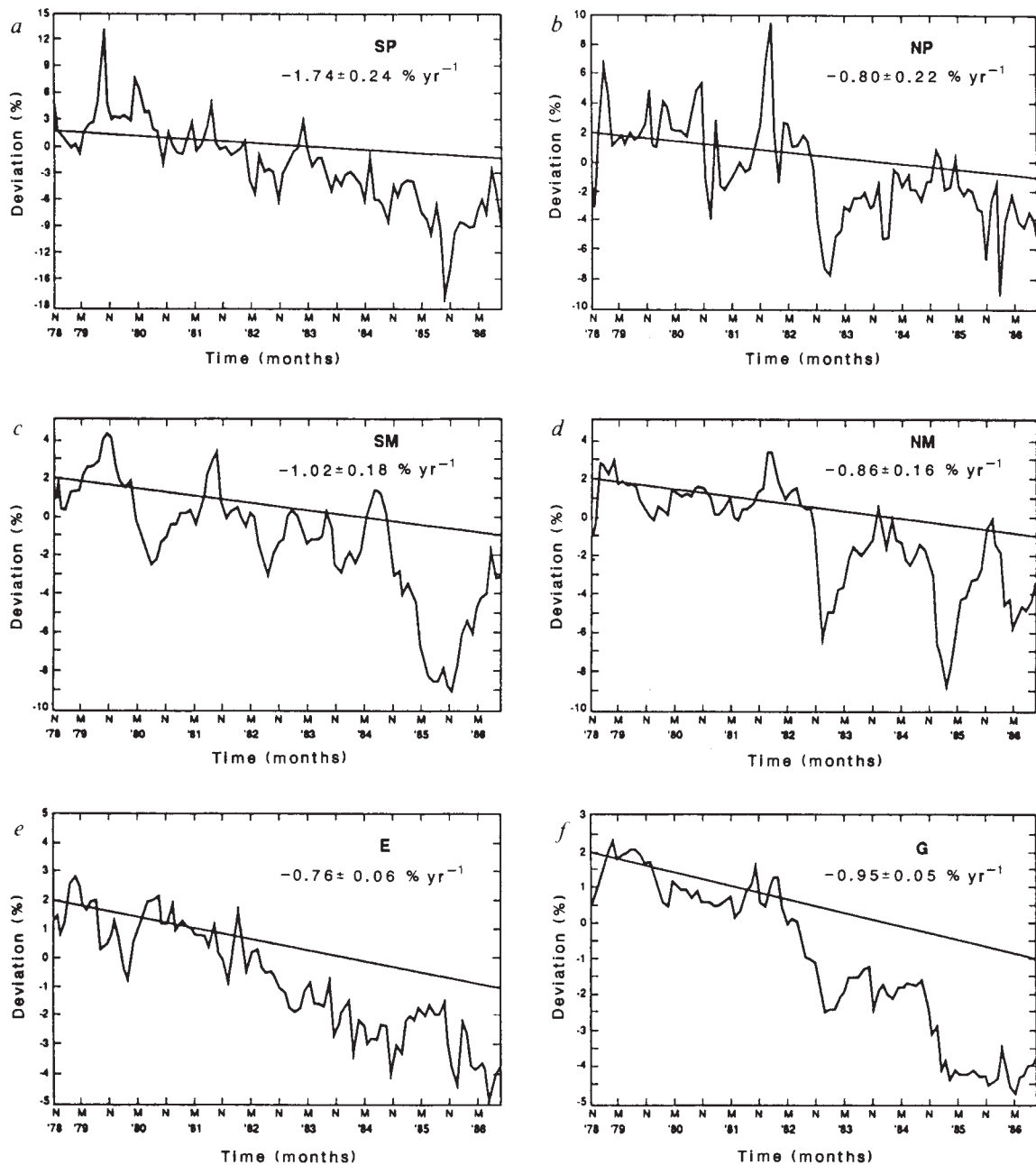


Fig. 2 Monthly deviations of total column ozone from six-year means for polar regions (81° – 55°), mid-latitude regions (55° – 25°), equatorial region (25° S– 25° N) and the globe (65° S– 65° N). Apparent drift of SBUV instrument relative to Dobson ground-based network is given by the straight line ($-0.38 \pm 0.13\% \text{ yr}^{-1}$). The uncorrected linear trends and 95% confidence limits are given for each region. To correct trends relative to Dobson network add $+0.38\% \text{ yr}^{-1}$.

($-0.43\% \text{ yr}^{-1}$) and SBUV ($-0.38\% \text{ yr}^{-1}$) instruments relative to the Dobson network⁹, that is, $+0.43\% \text{ yr}^{-1}$ and $+0.38\% \text{ yr}^{-1}$ would have to be added to the observed trends in total ozone from the respective Nimbus 4 and Nimbus 7 observations shown in Fig. 1. A tabulation of the drifts and the corresponding 95% confidence limits is given in Table 1.

The latitudinal trends shown in Fig. 1 indicate that total ozone was increasing at high latitudes and decreasing at mid- and tropical latitudes in the Southern Hemisphere from 1970–76. The rate of decrease is a minimum at 20° , which is not statistically significant, and is larger with increasing latitude in the Northern Hemisphere. Possible reasons for the apparent rate of decrease at high northern latitudes are discussed later. During recent years, 1978–86, the largest rates of decrease in total ozone were observed at high latitudes in the Southern Hemisphere. These rates decreased towards a minimum at the Equator the minimum

being still statistically significant at the 95% confidence level. In the Northern Hemisphere the poleward latitudinal gradient is less than the gradient in the Southern Hemisphere. The largest rates of decrease in total ozone were observed at 30° and 70° N in the Northern Hemisphere.

Area-weighted linear trends in global total ozone derived from the zonal mean data corrected for instrument drift relative to the Dobson network are $-0.26 \pm 0.17\% \text{ yr}^{-1}$ for 1970–76 and $-0.57 \pm 0.13\% \text{ yr}^{-1}$ for 1978–86. The 95% confidence limit of the trend uncertainty is obtained from the square root of the sum of the squares of the uncertainties in the drift correction and the global linear trend.

Non-seasonal changes. Monthly deviations of total ozone from six-year monthly means (November 1978–October 1984) are shown for five climate zones and globally for the period November 1978–October 1986 in Fig. 2. The climate zone means (polar,

Table 2 Annual deviations from means of column ozone shown in Fig. 2 for five climate regions and the globe

	South Polar	South temperate	Equatorial	North temperate	North polar	Globe
November 1978–October 1979	1.98	1.73	1.72	1.51	1.63	+1.68
November 1979–October 1980	4.02	0.63	0.93	0.76	2.76	+1.05
November 1980–October 1981	0.95	0.40	1.05	0.56	−0.24	+0.70
November 1981–October 1982	−1.48	−0.79	0.02	1.50	2.44	+0.23
November 1982–October 1983	−1.69	−0.81	−1.31	−3.06	−4.17	−1.78
November 1983–October 1984	−4.47	−1.15	−2.42	−1.28	−2.23	−1.88
November 1984–October 1985	−5.83	−5.84	−2.25	−4.96	−1.52	−3.91
November 1985–October 1986	−7.81	−5.45	−3.59	−3.44	−4.30	−4.15

Annual deviations are given in per cent.

Table 3 Large monthly deviations from monthly means of column ozone shown in Fig. 2 for five climate regions and the globe

	South polar	South temperate	Equatorial	North temperate	North polar	Globe
1978	November, +6.0					
1979	October, +13.4	November, +4.3			February, +6.5	April, +2.3
1980	May, +7.6				November, +5.3	
1981						
1982	June, −5.3				January, +9.5	
	November, −5.9					
1983	May, −6.0			January, −6.2	February, −7.6	January, −2.5
	October, −8.4					November, −2.4
1984			March, −3.4		January, −5.2	
			November, −3.9			
1985	July, −9.5	August, −8.6		March, −8.6	November, −6.6	March, −4.3
	October, −16.6	December, −9.0				November, −4.5
1986	July, −7.4		January, −4.3		February, −8.9	May, −4.7
	October, −9.2		August, −4.9			

Annual deviations given in per cent.

55°–81° N; temperate, 25°–55° N; equatorial, 25° S–25° N) are constructed from monthly latitudinal means of 10° width centred at 10° intervals from 80° S to 80° N. The global monthly deviations shown in Fig. 2 are derived from area-weighted zonal means which cover the latitudes 65° S–65° N. The drift of the SBUV instrument relative to the Dobson network¹⁰ is shown by the straight line in each of the panels in Fig. 2. Linear trends computed from a linear regression of monthly deviations from six-year means against time and the corresponding 95% confidence limits are given for each of the climate zones and globally in Fig. 2. The trends for these zones relative to the Dobson network can be obtained by adding 0.38% yr^{−1}. At present it has not been conclusively established that the sensitivity of the SBUV instrument is drifting at the rate given by the Dobson intercomparisons; consequently, the SBUV ozone data have not been corrected for drift against the Dobson network. Annual averages of the percentage monthly deviations from long-term monthly means shown in Fig. 2 are given in Table 2 for the five climate regions and the globe.

A close examination of the total ozone monthly deviations from long-term means shown for the five climate regions and the globe in Fig. 2 indicates that the interannual variability is less in the equatorial region than at high latitudes, and that the interannual variability exhibits a strong seasonal component which increases with latitude poleward from the Equator. In the polar regions, 55°–81°, the largest deviations often occur during the months following the polar night; January and February in the Northern Hemisphere and September–November in the Southern Hemisphere. Significant deviations are also often observed in both hemispheres during the months around the winter solstice, May–July in the Southern Hemisphere, and November and December in the Northern Hemisphere. At mid-latitudes, in the Northern Hemisphere, the largest

interannual changes in total ozone were observed in January, February, March and May. A characteristic feature of the equatorial region is an increasing rate of decrease after 1981.

A summary of large monthly deviations from long-term means shown in Fig. 2 is given in Table 3 for the five climate regions. Only deviations larger than 4% in the polar regions, 3% in the mid-latitude regions and 2% in the equatorial region and the global average are listed in Table 3.

The annual averages of the monthly percentage deviations of total ozone from seasonal means (Table 2) show several distinct features. An increase in the rate of decreasing total ozone did not appear simultaneously in the five climate regions and the globe. In the south polar region total ozone increased to a maximum value in 1980 and then began a nearly constant rate of decrease. In the southern mid-latitude region total ozone was decreasing at the rate of the apparent drift of the SBUV instrument relative to the Dobson network until 1984. Afterwards, in 1985 and 1986, total ozone decreased abruptly by ~4% below an extrapolated value from previous years. Total ozone in the equatorial region began to decrease at a greater rate after 1981. At northern mid- and polar-latitudes there is little change in total ozone during the years 1979–82, but a rapid decrease begins afterwards.

Seasonal trends. Patterns of the latitudinal dependence of the total ozone trends computed from deseasonalized monthly zonal mean data are shown in Fig. 3 for Nimbus 4 observations (1970–76) and Nimbus 7 observations (1978–86). The shading indicates regions where the confidence limit is <95% for the trend indicated by the contour line. It was assumed that a combination of a ratio of linear trend/standard error >1.6 for each element of 10° latitude and one month, when combined with smoothing of the contouring routine, was reasonable to specify the 95% confidence limits. The trends shown in Fig. 3

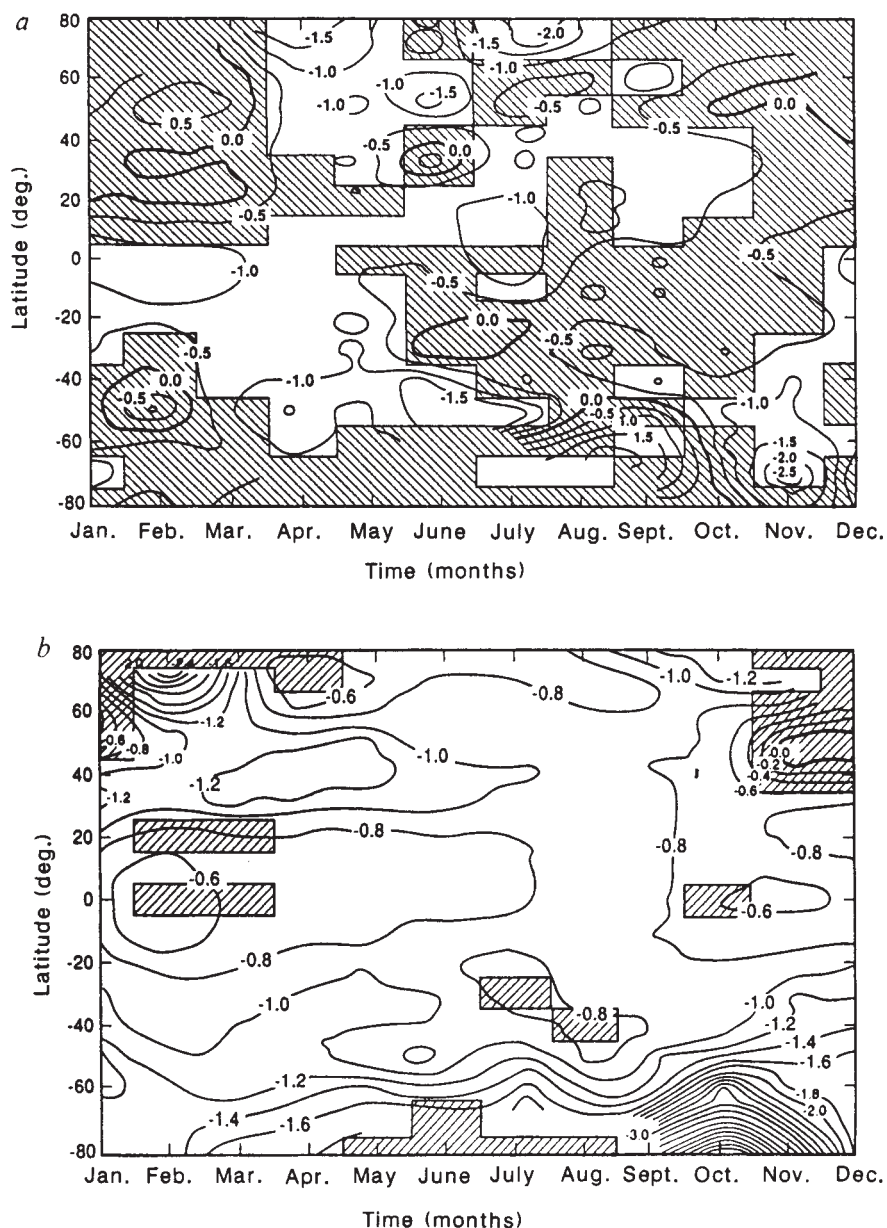


Fig. 3 Contours of total column ozone linear trends ($\% \text{ yr}^{-1}$) computed for each month and 10° latitude zone. Shaded areas indicate regions where the confidence limit is $<95\%$. No correction applied for possible instrument drift relative to Dobson network. *a*, Nimbus 4 monthly trends; *b*, Nimbus 7 monthly trends.

have not been corrected for instrument drift relative to the Dobson network (see Table 1). A correction for instrument drift will increase slightly the number of regions which are not statistically significant principally in regions of small trends, but it does not affect seriously the regions of large trends in either the Nimbus 4 or 7 data nor the conclusions.

The unshaded areas of the seasonal trends shown in Fig. 3 indicate that significant changes have occurred in the patterns and magnitude of the total ozone trends for 1970-76, Nimbus 4 observations, and 1978-86, Nimbus 7 observations. In general the patterns may be characterized by their behaviour at polar, temperate and tropical latitudes.

At northern polar latitudes ozone was decreasing at a rate of $1.5\text{--}2.0\% \text{ yr}^{-1}$ during the months April, May, July and August for 1970-76 and at rates from 2.4 to $1.0\% \text{ yr}^{-1}$ during February and March and from 1.0 to $1.2\% \text{ yr}^{-1}$ for September-November during 1978-86. Northern mid-latitude total ozone was decreasing at a rate of $\sim 1.0\% \text{ yr}^{-1}$ in April-June for 1970-76 and at a rate of $\sim 1.0\% \text{ yr}^{-1}$ during the months January-July for the period 1978-86. In the tropics, maximum rates of decreasing ozone are similar in magnitude during the Nimbus 4 and Nimbus 7 periods, $\sim 1\% \text{ yr}^{-1}$, but they differ in phase. The region of the

maximum rate of decreasing ozone in the tropics shifts from the months January-March for the Nimbus 4 period to the months July-September for 1978-86. At southern mid-latitudes similar rates of ozone depletion, $\sim 1.0\% \text{ yr}^{-1}$, are observed during the months March-May during both the Nimbus 4 and 7 periods. Trends in total ozone in the southern polar regions show large differences between the two periods. During the period 1970-76 ozone was increasing during July and August, but decreasing at a rate of $2.5\% \text{ yr}^{-1}$ during November with the net annual effect (Fig. 1) of being slightly negative with respect to the tropics. During the period 1978-86 ozone at southern polar latitudes shows strong poleward gradients in the rate of decrease, reaching values $>1.6\% \text{ yr}^{-1}$ during the months March-September and $\sim 4.8\% \text{ yr}^{-1}$ during October.

Spatial patterns of trends. Additional information on the importance of various physical processes which are producing ozone changes on a global scale may be obtained from the geographical distribution of derived linear trends from 1978 to 1986. These changes shown in Fig. 4, are derived from a linear trend analysis of monthly deviations from six-year monthly means computed from monthly average global synoptic analyses on the standard National Meteorological Center (NMC) 65×65 rectangular

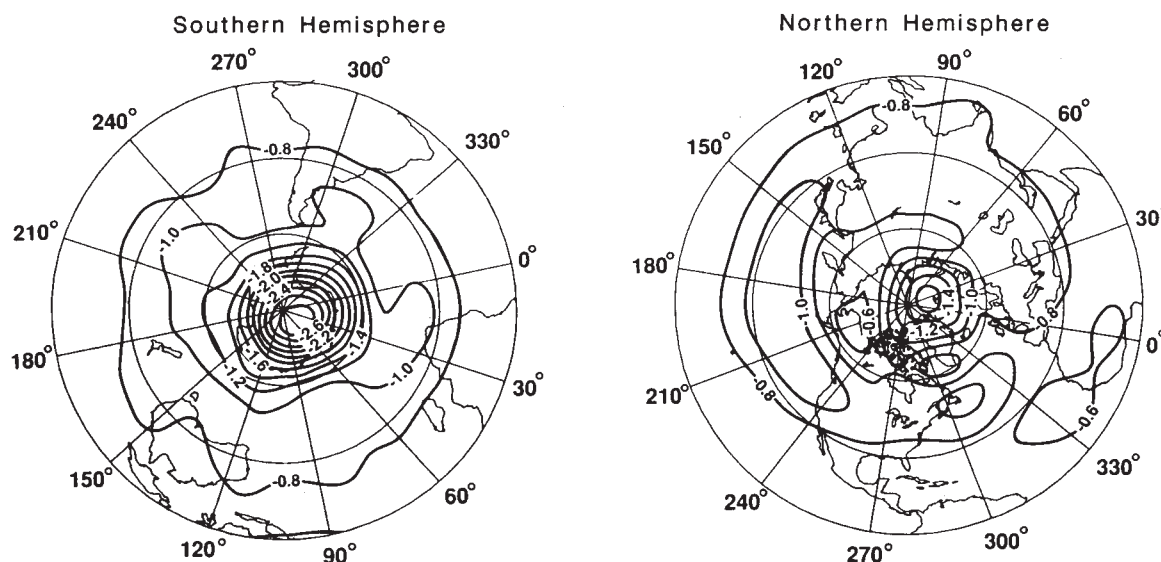


Fig. 4 Contours of linear trends computed from monthly deviations over the period 1978–86 from an average of six-year monthly means on a 65×65 NMC grid. No correction applied for possible instrument drift relative to Dobson network.

array on polar stereographic projections of the Northern and Southern Hemispheres. Since the maximum poleward latitude is 81° and observations are used only to within 2° of the terminator, values are interpolated across the pole. The standard NMC interpolation scheme uses a combination of gradients and wave patterns in existing data and a $1/r^2$ weighting as it extends outwards searching for other data points. Consequently the interpolation scheme does not create features in the analysis which do not already exist in regions of data, but weakens the patterns poleward. These data, uncorrected for instrument drift with respect to the Dobson network, are contained on the SBUV Contours Tape¹¹ and are available from the National Space Science Data Center.

In the Northern Hemisphere the largest rates of ozone change, $\sim -1.6 \pm 0.4\% \text{ yr}^{-1}$, are observed in the Arctic, in a region offset from the pole and almost centred over the island of Spitzbergen, Norway and encompassing the northern tip of Greenland, Norway, Sweden, Finland and the USSR. In contrast, there is a region of only slightly decreasing ozone, $-0.6 \pm 0.4\% \text{ yr}^{-1}$, over southern Alaska. At northern mid-latitudes the region of largest rates of decrease of $1.0 \pm 0.1\% \text{ yr}^{-1}$, is located over the Pacific Ocean between Japan and the western United States. In the northern tropics the largest rates of decreasing ozone ($-0.8 \pm 0.2\% \text{ yr}^{-1}$) are observed in the vicinity of Malaysia and the southern regions of India and Thailand.

The most notable feature in the Southern Hemisphere is the Antarctic region of ozone change of $-2.6 \pm 0.4\% \text{ yr}^{-1}$ which is offset slightly from the pole along the longitude meridian of 30° E as it is in the Northern Hemisphere. The location of this region coincides with the region of the October minimum in total ozone¹². The southern mid-latitudes are characterized by three regions of high rates of ozone depletion, $-1.0 \pm 0.2\% \text{ yr}^{-1}$, that extends from $\sim 30^\circ \text{ S}$ – 40° S equatorwards to almost 30° and nearly centred along longitudes 30° , 220° and 310° .

At polar latitudes the patterns of the ozone trends in the Northern Hemisphere suggest a wave number 1 configuration whereas in the Southern Hemisphere they are nearly zonal. The distributions of total column ozone and 30- and 70-mbar temperatures during winter at polar latitudes in both hemispheres are similar to the patterns of ozone depletion. (Monthly mean maps of SBUV total ozones 30- and 70-mbar temperatures are available from the Climate Analyses Center of the National

Oceanic and Atmospheric Administration (NOAA).) The regions of large rates of ozone depletion coincide with regions of low temperature and low total ozone. In the region 54° – 84° S ozone observations using TOMS during October have been shown to be well correlated with 70-mbar temperatures for the years 1979–85¹³. Most of these similar changes in total ozone during October appear to be directly correlated with declining temperatures in the stratosphere which they interpret as evidence of a secular dynamic change in the stratosphere which is changing both total ozone and temperature.

Discussion and conclusions

Apparent drift of SBUV. The question of satellite instrument drift relative to the Dobson network is important for comparisons of trends obtained during the periods of 1970–86 and for future years. While a complete discussion of the differences between measurements of total ozone with the ground-based Dobson instruments and the satellite-borne SBUV instruments and an intercomparison of the measurements is beyond the scope of this paper, several important aspects of the apparent drift of the SBUV type instruments relative to the Dobson network are described. The SBUV and Dobson instruments do not use the same wavelengths to determine total ozone. The Dobson AD pair measurement of total ozone is sensitive to errors due to atmospheric SO_2 and NO_2 ¹⁴. The SBUV instrument, by virtue of the height dependence of the weighting functions of the total ozone wavelengths, possesses only one-half of the Dobson instruments' sensitivity to changes in tropospheric ozone¹⁵. However, based upon our current knowledge of increasing levels of tropospheric ozone and atmospheric SO_2 (ref. 14) this does not appear to be able to account for the observed rate of drift.

A study¹⁰ of selected stations of the Dobson network gives $-0.38 \pm 0.13\% \text{ yr}^{-1}$ as the drift of the SBUV instrument relative to the Dobson network. Analyses of the instrument characterization which are independent of the Dobson network indicate an uncertainty in the instrument calibration of $0.15\% \text{ yr}^{-1}$ at the 95% confidence level¹⁶.

A recent unpublished update of the apparent drift of the SBUV instrument relative to the Dobson network over eight years indicated that the drift could be characterized by a linear fit of $-0.44 \pm 0.12\% \text{ yr}^{-1}$ or by two linear fits of $-0.29 \pm 0.18\%$

yr⁻¹ up to July 1982 and $-0.57 \pm 0.24\%$ yr⁻¹ afterwards (C. Wellemeyer and P. K. Bhartia, personal communication).

An approximate rate of increase of tropospheric ozone of 1% yr⁻¹ could account for only $\sim 0.05\%$ yr⁻¹ of the observed rate of drift between the two systems which is nearly an order of magnitude larger. The Dobson instrument sensitivity of the AD double wavelength pair to SO₂¹⁴ would require a rate of increase in column SO₂ of $\sim 0.8 \times 10^{-4}$ atm cm⁻¹ yr⁻¹ to account for the apparent drift of the SBUV instrument which seems unreasonable in terms of average background levels of SO₂.

Additional information on the sources of the drift of the SBUV instrument relative to the Dobson network can be obtained by comparing total ozone derived from the SBUV standard wavelength pairs 312.5 nm/331.2 nm (A-pair) and 317.5 nm/339.8 nm (B-pair) with that derived from the wavelength pairs 305.8 nm/312.5 nm (D-pair) and 305.8 nm/331.2 nm (D'-pair). The shortest wavelength of D, D' pairs is very close to the shortest wavelengths (305.5 nm) of the Dobson AD double wavelength pair. Preliminary results of a study of the drifts in total ozone derived with the different wavelength pairs within 10° latitude zones centred at 20° N, 0° and 15° S indicated significant differences between the period before and after the eruptions of El Chichon in April 1982 (S. Taylor, C. Wellemeyer and P. K. Bhartia, personal communication).

Before the eruptions of El Chichon there were no differences in ozone derived with the A-pair and B-pair for the three latitude zones. The approximate drift between A-pair ozone minus D-pair ozone was -0.23% yr⁻¹ at 20° N and nearly 0 at 0° and 15° S. After the eruptions of El Chichon significant drifts were observed in the A-pair and B-pair ozone differences and the A-pair and D-pair ozone differences. While the drift in A-pair and D-pair ozone differences was about $+0.20\%$ yr⁻¹ in each of the three latitude zones, the drift with A-pair and D-pair ozone differences was $\sim -0.23\%$ yr⁻¹ (same as the pre-El Chichon period) at 20° N, -0.61% yr⁻¹ at the Equator and -0.50% yr⁻¹ at 15° S. These preliminary observations indicate that something other than a wavelength-dependent drift in the instrument sensitivity is required to explain these drifts in the differences in total ozone derived from different wavelength pairs at different latitudes before and after the eruptions of El Chichon in April 1982. Errors in the determination of instrument sensitivity with time would not be expected to give biases in total ozone which are latitude dependent and whose latitude dependency changes following the eruption of El Chichon.

The magnitude of the SBUV instrument drift relative to the Dobson network can also be assessed by comparing deviations from long-term means for the climate zones and globally. This would tend to minimize regional effects associated with the clustering of Dobson stations at mid- and high latitudes in the Northern Hemisphere. Averaging and smoothing the SBUV total ozone data, as was done for the ground-based data shown in Fig. 5¹⁷, yielded rates of SBUV instrument drift of -0.54% yr⁻¹ in the south temperate zone, -0.32% yr⁻¹ in the equatorial zone, -0.34% yr⁻¹ in the north temperate zone and -0.17% yr⁻¹ globally for the period between the summer of 1979 and the winter of 1986–87 in the Northern Hemisphere. The differences for the period between 1979 and 1986–87 for the SBUV observations and analyses of the ground-based observations were -9.5% and -6.0% respectively for the south temperate region, -4.6% and -2.5% for the equatorial region, -4.2% and -2.0% for the north temperate region and -6.1% and -5.0% for the globe. A combination of effects associated with the polar regions, weighting of the zones and the small number of stations in the Southern Hemisphere probably contributed to the differences between the calculated linear rates of drift between the equatorial and north temperate regions and the globe. A mean of the equatorial and global values indicates an apparent drift of the SBUV instrument relative to the ground-based network of $\sim -0.25\%$ yr⁻¹ which is less than that stated previously.

The occurrence of polar stratospheric cloud in the coldest regions of the Antarctic have been linked²⁹ to the depletion of the stratospheric ozone in the Antarctic austral spring. Measurements of polar stratospheric clouds in the Antarctic and Arctic suggest that the integrated optical depths are <0.01 (refs 18–20). It has been shown²¹ that even clouds with an optical depth of one have a negligible effect on the retrieval of total ozone by SBUV type instruments. The problem of cloud-induced biases in the measurements of total ozone by the SBUV, BUV and TOMS instruments has been examined extensively and no evidence for a bias, much less a time-dependent one has been found.

Regional and global changes

The Nimbus 4 BUV observations are of limited usefulness for determining global changes during the period 1970–77 because of tape recorder and power problems on the spacecraft. Even during the first two years when the coverage was best, data are missing most of the time between longitudes of 120–180° E as discussed previously. These missing data could bias linear trends computed from latitudinal zonal means if significant longitudinal asymmetries are present in the year-to-year changes. Because the linear trends shown in Fig. 4 for the Northern Hemisphere show considerable asymmetry one should not conclude from Fig. 1 that an Arctic sink for total ozone was stronger in the 1970s than after 1978. The symmetry of the linear trends in the Southern Hemisphere shown in Fig. 4 suggests that the rate of ozone change for southern high latitudes (see Fig. 1) for 1970–76 is real and the rate of ozone decrease in the Antarctic has been accelerating during the period 1970–86. These changes are consistent with the reported decreases of October monthly means of total ozone at Halley Bay which began to decrease rapidly after 1976³.

Before mid-1982 the monthly total column ozone deviations from the mean shown in Fig. 2 for northern mid-latitudes and the polar region indicated no obvious downward trend such as can be seen for the equatorial latitudes and those towards Antarctica. After 1982 mid- and high-latitude ozone began to decrease substantially in the Northern Hemisphere. Following the eruptions of El Chichon in April 1982 and the extremely strong El Niño–Southern Oscillation event during the Northern Hemisphere winter of 1982–83, decreases in total ozone of $\sim 6.5\%$ were observed in the mid-latitude region and $\sim 8\%$ in the north polar region. Infrared measurements with the TOVS instrument (at HIRS) and ground-based observations indicated a 5–7% decrease in total ozone in the North American region and north temperate zone yielding the lowest values on record since about 1958²².

During the period following these primarily Northern Hemisphere events the ozone content of the atmosphere from 65° S–65° N declined by $\sim 2.5\%$ (Fig. 2). Because the ozone minimum observed in the Northern Hemisphere during the winter of 1982–83 was also a period of significant climatic circulation anomalies resulting from the intense El Niño–Southern Oscillation of 1982–83 it might be reasonable to assume that the ozone minimum was mainly the result of changes in ozone transport. Some of these anomalies observed were the lowest values yet recorded²³. The assumption that these climatic circulation anomalies were responsible for the deep Northern Hemisphere winter ozone minimum is not supported by analysis of observations over the past 26 years of the relationship between equatorial sea surface temperatures (a measure of the Southern Oscillation Index) and north temperate total ozone²².

A second period of greatly reduced total ozone at mid-latitudes is shown in Fig. 2. Reductions of $\sim 9\%$ from a long-term mean were observed in the mid-latitude region of the Northern Hemisphere during the winter of 1984–85 and in the mid-latitude region of the Southern Hemisphere during the winter of 1985. These winter perturbations reduced the total ozone in the region 65° S–65° N by $\sim 2.5\%$ and significantly increased the 1978–86

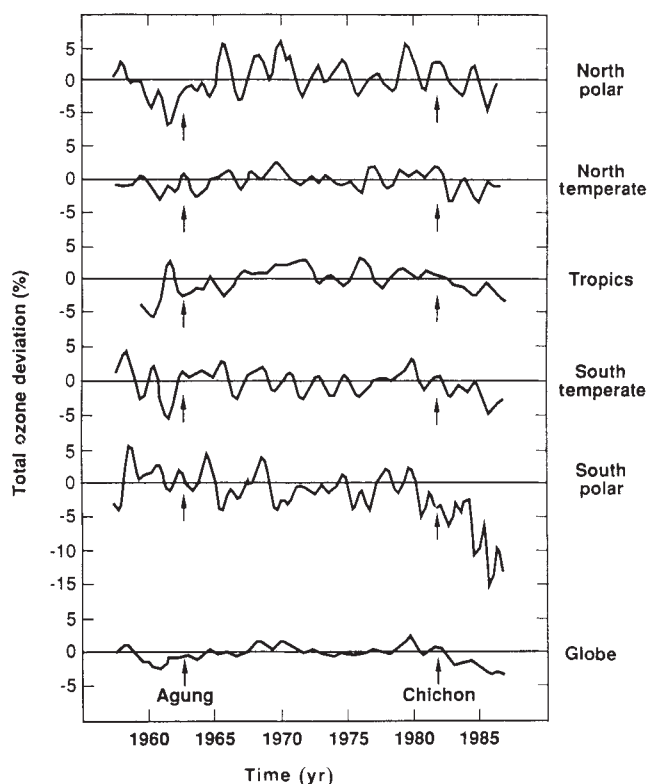


Fig. 5 Climate zone and global average of seasonal deviations from a long-term mean of Dobson network total ozone data (courtesy of J. K. Angell).

rate of decrease shown in Fig. 1 at mid-latitudes.

An area weighting of the linear trends shown in Fig. 1 in terms of ozone amount indicates that the latitude regions poleward of 65° were not the major contributors to the decline in the global budget of total ozone. These contributions to uncorrected and corrected global trends are 11% and 15% respectively for zones 70° S and 80° S and 6% for both zones 70° N and 80° N.

The monthly and latitudinal patterns of the ozone trends shown in Fig. 3 are suggestive of several physical mechanisms associated with increasing rates of ozone depletion in the late winter-early spring period at high latitudes in both hemispheres and between April and September in the tropics. The strong interannual variability observed at high latitudes in Fig. 2 is suggestive of dynamical processes which are strong during late winter-early spring. Recent two-dimensional model predictions²⁴ for changes in chlorine levels (from 1 to 2.7 p.p.b.) and bromine levels (from 15 to 30 p.p.t.) predict maximum high-latitude depletions of total ozone of 4% during February and March in the Northern Hemisphere and 12% during October in the Southern Hemisphere for the change from the pre-perturbed atmosphere to that in 1985. While the high-latitude patterns are similar to both the observed and modelled trends, the observed maximum rates of depletion of $\sim 2.4\% \text{ yr}^{-1}$ in the Northern Hemisphere and $4.8\% \text{ yr}^{-1}$ in the Southern Hemisphere, which for eight years yielded reductions of 19% and 38%, are not, although the ratio is similar. Work which will be reported in a separate publication shows that the largest long-term changes are observed above 10 mb and below 30 mb.

The total ozone monthly deviations from long-term means shown in Fig. 2 for the northern mid-, equatorial, southern mid-latitude regions and the area-weighted global averages (65° S– 65° N) derived from SBUV observations and uncorrected for possible instrument drift are in reasonable agreement with

the seasonal deviations given in Fig. 5 which are derived from an analysis of observations of the ground-based network¹⁷. A correction for instrument drift against the Dobson network¹⁰ is given by a straight line in each panel. Even if the apparent drift of the SBUV instrument relative to the Dobson ground-based network shown in Fig. 2 were real it is apparent that there have been step-wise decreases in the global budget of total ozone (65° S– 65° N). These decreases are due mainly to large seasonally dependent decreases during the winter-spring period in mid-latitude regions in the Northern and Southern hemispheres and an accelerated rate of decrease in the equatorial region after 1981. It is important to note that the maximum contribution from the regions poleward of 65° in both hemispheres to the observed decreases in global total ozone is probably no more than about 20%. Based upon the most recent update of the analysis given in Fig. 5, the decline from 1979 to 1986 and the low values after 1984 are the largest ever recorded by the ground-based network¹⁷ even though the period of the beginning of SBUV observations in late 1978 appears to have been a record high (Fig. 5) for global total ozone.

The geographical patterns of total column ozone trends shown in Fig. 4 for 1978–86 indicate that the regions of large rates of decrease are the Antarctic, the Arctic and a mid-latitude region in the Northern Hemisphere extending from $\sim 30^\circ$ latitude (near the tip of Japan) and which extends northward and eastward. This region, when examined with greater resolution than shown in Fig. 4, appears to be part of a clockwise spiral extending southward from the Arctic region of maximum ozone depletion over Spitzbergen Island, Norway. The two regions showing a minimum in the rates of decreasing ozone are located over northern Alaska and westward into the Atlantic Ocean from the western most part of the African continent.

The NMC monthly means maps of global synoptic analyses of total ozone¹¹ that were used for the analyses shown in Fig. 4 indicate that the location of the Arctic ozone region of enhanced depletion coincides with the region of low ozone extending polewards from northern Europe. This region of minimum high-latitude ozone is observed during each February from 1979–86. Similarly, high-latitude regions of high ozone appear to be associated with regions of smaller rates of decreases in ozone. A similar behaviour is observed during winter at south polar latitudes¹².

The location of the maximum rate of ozone decrease in the Arctic coincides with the average location of regions of lowest 30-mbar temperatures for the months October–February from 1978–86. The 30-mbar temperature data were taken from analyses available in the Beilage zur Berliner Wetterkarte from the Free University of Berlin. The region of coldest average autumn and winter temperatures is 77° N latitude and 29° E longitude which lies in the centre of the Arctic region of enhanced ozone depletion. The region of warmest average autumn and winter temperatures is 56° N latitude and 158° E longitude which is near the same latitude but $\sim 50^\circ$ west of the region of slightly increasing ozone at high latitudes. This suggests that important factors contributing to the growth of the Arctic ozone region of enhanced depletion could be the isolation from mixing with warmer air and higher ozone concentrations observed at lower latitudes and low temperatures or changes in the diabatic circulation.

Another region of decreasing ozone is located in the western Pacific and the Indian Ocean between 60° – 150° E longitude slightly north of the Equator which coincides with the region proposed for the injection of water vapour into the stratosphere and the coldest stratospheric temperatures²⁵ outside the Antarctic region.

The latitudinal distribution of the seasonal trends observed in total ozone for 1978–86 indicates that tropical ozone is decreasing most rapidly in a region north of the Equator during the months April–September. During these months, lower stratospheric temperatures are especially low over the Bay of

Bengal, India and Malaysia²⁵. An extensive review of stratospheric-tropospheric exchange processes which could affect the observed decreases in total column ozone in the tropics has been published²⁶, however, a discussion of these is beyond the scope of this paper.

A brief report²⁷ on a recent analysis of global trends in total ozone derived from the Nimbus 7 total ozone mapping spectrometer (TOMS) for the years 1979–86 is in reasonable agreement with the results from the Nimbus 7 SBUV instrument for similar types of analyses. This is not surprising as both instruments use the SBUV diffuser plate for measurements of solar spectral irradiance but each views significantly different areas of the diffuser plate.

The analysis of seasonal TOMS data yields linear trends which are uncorrected for drift against the Dobson network of -1.0% yr⁻¹ while the SBUV analysis of deseasonalized data also uncorrected for drift gives $-0.97 \pm 0.05\%$ yr⁻¹. It has also been suggested²⁷ that the apparent drift against the Dobson network may arise partially from geophysical processes. The analysis of TOMS total ozone data indicate that the largest rates of decreasing ozone in both hemispheres are associated with regions over east Antarctica and north of Scandinavia. This agrees with the results obtained from the analyses of SBUV total ozone data. Both analyses show that the decreases in global total ozone are not solely the result of decreases in the Antarctic during spring.

None of the current theories or model predictions for either natural or anthropogenic sources of stratospheric ozone depletion appear capable of explaining the large decreases observed in total column ozone by both the satellite SBUV and Dobson network instruments. The magnitude of the changes suggests that a major climatic change is taking place in the stratosphere resulting from a combination of photochemical and dynamical processes. The very rapid decline toward the minimum of solar cycle 21 of ultraviolet solar spectral irradiance after mid-1981²⁸, the eruptions of El Chichon in April 1982, and the very strong El Niño–Southern Oscillation event during the 1982–83 winter and changed levels of anthropogenic constituents should be considered as factors in determining why the response of stratospheric ozone to natural perturbations appears to be different during the period 1978–85 than it was in the past. Additional information which can contribute to the evaluation of physical processes that could produce the observed changes in total column ozone can be obtained from similar analyses of the

long-term changes in the vertical distribution of stratospheric ozone. The analyses that have been performed indicate that the largest long-term changes are observed at altitudes between 10 mbar and 30 mbar.

In conclusion analyses of non-seasonal changes in global total column ozone derived from satellite observations during the period 1970–86, when compared with analyses derived from the Dobson network, indicate that the largest global decrease ever recorded during the period 1959–86 of $\sim 5\%$, occurred between 1978–86. Linear trend patterns of deseasonalized ozone data show that the high latitudes in both hemispheres are regions of the most rapid ozone decreases, however, latitudes poleward of 65° contributed no more than 20% of the average global decrease during the period 1979–86. The maximum high-latitude rates of change are observed during February in the Northern Hemisphere ($\sim -2.4\%$ yr⁻¹) and during October in the Southern Hemisphere ($\sim -4.8\%$ yr⁻¹). Geographical distributions of linear ozone trends for the period 1978–86 confirm the existence of a region of rapidly decreasing ozone over the Antarctic of -2.6% yr⁻¹ and indicate the existence of a similar Arctic region of -1.6% yr⁻¹, centred over Spitzbergen Island, Norway. These polar regions of maximum rates of ozone depletion coincide with the regions of lowest 30-mbar temperatures. The existence of a tropical region of enhanced ozone depletion is reported which appears to coincide with the location of especially cold sub-polar temperatures, strong vertical motions and a suggested source of stratospheric water vapour. Analyses indicate consistently that the largest rates of ozone decreases are associated with climatological regions of lowest total ozone and coldest temperatures. Beginning in 1983, the mid-latitudes have become additional sources of ozone decreases which have provided significant contributions to the continuing decline in the global ozone content. Recent analyses of TOMS data also confirm the global decrease in total ozone observed with the SBUV instrument and the Dobson network and an Arctic region of enhanced ozone depletion.

The observed enhanced rates of ozone decreases during the months of winter and early spring, the coincidence of regions of large rates of ozone decreases with regions of lowest temperatures and the two step-like decreases in the global budget after 1981 suggest that much photochemical, radiative and dynamical modelling needs to be done to determine the importance of various processes that may have produced the global-scale changes observed between 1978–86.

Received 2 February 1987; accepted 1 March 1988.

- Johnston, H. S. *Science* **173**, 517–522 (1971).
- Molina, M. J. & Rowland, F. S. *Nature* **249**, 810–812 (1974).
- Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
- Dobson, G. M. B. *Ann. Int. Geophysical Year V Part I*, 46–81 (Pergamon, New York, 1957).
- Heath, D. F., Mateer, C. L. & Krueger, A. J. *Pure appl. Geophys.* **106–108**, 1239–1253 (1973).
- Heath, D. F., Krueger, A. J., Roeder, H. A. & Henderson, B. D. *Opt. Engng* **14**, 323–331 (1975).
- Klenk, K. F., Bhartia, P. K., Fleig, A. J. & Mateer, C. L. *Fifth Conf. on Atmospheric Radiation* Baltimore, Maryland, 103–106 (American Meteorological Society, Boston, 1983).
- Bhartia, P. K., Klenk, K. F., Gordon, D. & Fleig, A. J. *Fifth Conf. on Atmospheric Radiation* Baltimore, Maryland, 115–117 (American Meteorological Society, Boston, 1983).
- Labitzke, K. & Miller, A. J. in *Atmospheric Ozone 1985* Vol. III, Ch. 14 (World Meteorological Organization, Geneva, 1986).
- Fleig, A. J., Bhartia, P. K. & Silberstein, D. S. *Geophys. Res. Lett.* **13**, 1359–1362 (1986).
- Fleig, A. J. *et al. Users Guide for SBUV/TOMS Ozone Derivative Products* (NASA Publ. No. 1116) (1984).
- Stolarski, R. S. *et al. Nature* **322**, 808–811 (1986).
- Schoeberl, M. R. & Krueger, A. J. *Geophys. Res. Lett.* **13**, 1217–1220 (1986).
- Komhyr, W. D. & Evans, R. D. *Geophys. Res. Lett.* **7**, 159–160 (1980).
- Klenk, K. F. *et al. J. appl. Meteor.* **21**, 1672–1684 (1982).
- Cebula, R. P., Park, H. & Heath, D. F. *H.J. atmos. oceanic Technol.* (in the press).
- Angell, J. K. *Mon. Weath. Rev.* (in the press).
- McCormick, M. P. & Larsen, J. C. *Geophys. Res. Lett.* **13**, 1280–1283 (1986).
- Kent, G. S., Poole, L. R. & McCormick, M. P. *J. atmos. Sci.* **43**, 2149–2160 (1986).
- McCormick, M. P. & Trepte, C. R. *J. geophys. Res.* **92**, 4297–4306 (1987).
- Dave, J. V. *J. atmos. Sci.* **35**, 899–911 (1978).
- Angell, J. K., Korshover, J. & Planet, W. G. *Mon. Weath. Rev.* **113**, 641–646 (1985).
- Quiroz, R. S. *Mon. Weath. Rev.* **111**, 1685–1706 (1983).
- Isaksen, I. S. A. & Stordal, F. *Geophys. Res. Lett.* **13**, 1327–1330 (1986).
- Newell, R. E. & Gould-Stewart, S. *J. atmos. Sci.* **38**, 2789–2796 (1981).
- Tuck, A. F. in *Atmospheric Ozone 1985* Vol. I, Ch. 5, 151–240 (World Meteorological Organization, Geneva, 1986).
- Bowman, K. P. *Science* **239**, 48–50 (1988).
- Heath, D. F. & Schlesinger, B. M. *J. geophys. Res.* **91**, 8672–8682 (1986).
- Solomon, S., Garcia, R., Rowland, F. S. & Wuebbles, D. E. *Nature* **321**, 755–758 (1986).

Possible demographic consequences of AIDS in developing countries

R. M. Anderson, R. M. May* & A. R. McLean

Parasite Epidemiology Research Group, Department of Pure and Applied Biology, Imperial College, London University, London SW7 2BB, UK

* Biology Department, Princeton University, Princeton, New Jersey 08540, USA

Simple mathematical models of the transmission dynamics of HIV that incorporate demographic and epidemiological processes to assess the potential impact of AIDS on human population growth and structure in developing countries suggest that AIDS is capable of changing population growth rates from positive to negative values over timescales of a few decades. The disease is predicted to have little if any impact on the dependency ratio of a population, defined as the number of children below age 15 years and elderly people over 64 years, divided by the number of adults between 15 to 64 years.

A LARGE number of cases of acquired immune deficiency syndrome (AIDS) are now being reported to the World Health Organization (WHO)—in October 1987, 123 countries reported at least one case¹—and evidence is accumulating of the degree to which the aetiological agent of the disease (Fig. 1), the human immunodeficiency virus (HIV), has penetrated heterosexual communities in developing countries (particularly in Sub-Saharan Africa and South America²⁻⁷). This has prompted governments and international agencies to give urgent consideration to the degree to which the pandemic of AIDS will influence the demography and socio-economic development of large areas of the world in the coming decades.

Scientific assessment of this problem is complicated by our limited understanding of the epidemiology of the infection (HIV) and the disease (AIDS)³. Uncertainties include: the fraction of those infected who will proceed to develop AIDS and the time scale of this progression⁵⁻⁹; the likelihood of vertical transmission from infected mother to child *in utero*, during childbirth, and by breast feeding¹⁰⁻¹⁴; the pathogenicities of human retroviruses other than HIV-1 (such as HIV-2) that are currently spreading in certain countries¹⁵; the importance of cofactors such as genital ulcers in heterosexual transmission¹⁶; the degree to which the infectiousness of infected patients changes throughout incubation of the disease^{17,18}; the probabilities of horizontal transmission from male to female and vice versa^{19,20}; and patterns of sexual behaviour in defined communities²¹⁻²³. The long and variable incubation period of AIDS means that epidemiological knowledge will only accumulate slowly from long-term studies of infection and disease incidence, the rise in the former preceding that in the latter by many years^{3,11}. Current estimates of the mean period between HIV infection and development of AIDS, derived from transfusion-associated cases in the United States, are 8-9 years²⁴, although the mean may be lower in developing countries where people are exposed more frequently, and to a larger range of infectious agents, than in the developed world^{3,11,25}. A combination of social and political sensitivities, and the more practical difficulties in accurate diagnosis and reporting of infection and disease in poor countries, has hindered the study of the spread of infection^{3,11,25,26} in these parts of the world.

Quantitative data on the epidemiological processes outlined above, plus data on age-specific fertility and mortality, is essential for precise predictive work on the demographic impact of AIDS. The need to assess the magnitude of the problem however, is urgent to facilitate the long-term planning of governments and international aid agencies. It therefore seems sensible to study mathematical models that combine the demography of populations having positive net growth rates with the presently known epidemiological characteristics of the heterosexual transmission of HIV-1. In this paper we develop simple models with the aim of obtaining a crude understanding of how AIDS deaths

might affect demographic patterns, of the timescales of such effects and of the influence of demographic and epidemiological parameters on population sizes. Mathematical study of these deliberately simplified models is a preliminary to future numerical exploration of much more complicated and realistic models possible once data accumulates. The qualitative understanding that simple models can provide make it easier to recognize what needs to be measured, together with the appropriate timescales and parameter ranges in subsequent and more elaborate studies^{19,26,27}.

Definition of a basic model

A theoretical framework for exploring how infectious diseases may affect animal population sizes has been developed over the past decade²⁸⁻³¹. More recently, a number of mathematical studies have combined demographic and epidemiological considerations to examine the impact, transmission, and control by mass vaccination, of directly-transmitted childhood infections such as measles in developing countries^{32,33}. A mathematical framework for the study of the transmission dynamics of HIV has begun to be developed^{19,26,34-36} but as yet it does not address the question of demographic impact.

We begin by considering a total population $N(t)$ at time t , subdivided into $X(t)$ susceptibles and $Y(t)$ infecteds³⁷ (assumed to be infectious), all of whom develop AIDS on some characteristic long timescale (a growing body of evidence suggests that a very high fraction of those infected will eventually develop symptoms of the disease^{38,39}, so in this basic model the fraction, f , of infecteds who develop AIDS is assumed to be one). For simplicity, we do not consider a separate class for AIDS patients as the average incubation period ($1/(\alpha + \mu)$) appears long in relation to the life expectancy of an AIDS patient ($1/(\alpha + \mu) > \text{nine years}$ while life expectancy with AIDS is about one year). In this calculation α is the disease-related death rate and deaths from all other causes occur at a constant rate μ (type II survival, giving an exponential age distribution which is a reasonable approximation for many developing countries). These assumptions yield the following pair of differential equations for $N(t)$ and $Y(t)$:

$$dN/dt = N[(\nu - \mu) - (\alpha + (1 - \varepsilon)\nu)(Y/N)] \quad (1)$$

$$dY/dt = Y[(\beta c - \mu - \alpha) - \beta c(Y/N)] \quad (2)$$

We define the net birth rate of the community B as:

$$B = \nu[N - (1 - \varepsilon)Y] \quad (3)$$

Here ν is the per capita birth rate (females per female, or equivalently offspring per capita for a 1:1 sex ratio) in the absence of infection; ε is the fraction of all offspring born to infected mothers who survive and $(1 - \varepsilon)$ is the fraction who acquire infection by vertical transmission and die rapidly from

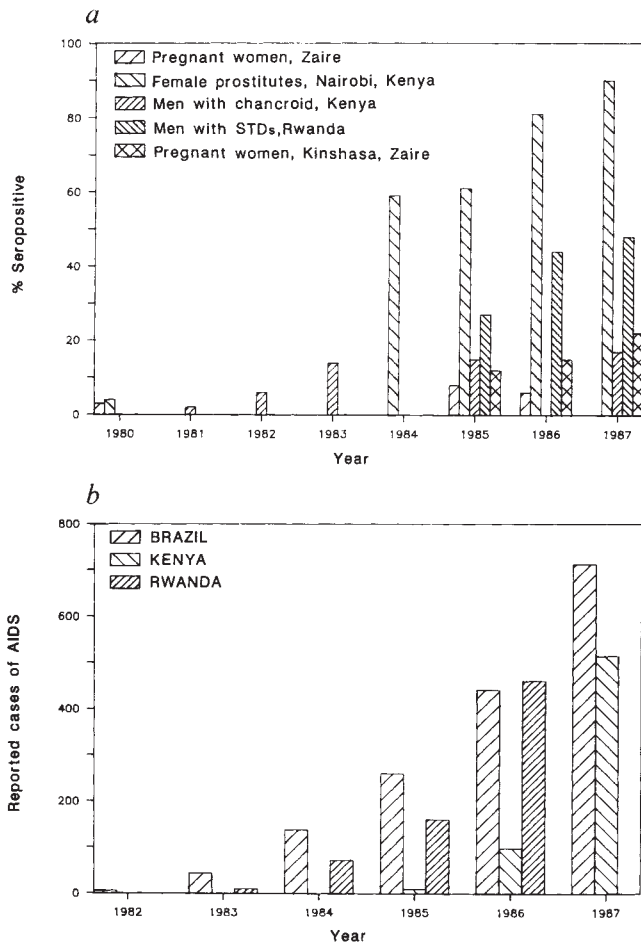


Fig. 1 *a*, Longitudinal changes in the proportion of people in various at-risk groups in Africa, who have antibodies to HIV-1 antigens (seropositive) (data sources refs 11, 48, 50–52). □, Pregnant women in Zaire; ▨, Female prostitutes in Nairobi, Kenya; ▩, Men with chancroid, Kenya; ▤, Men with STD infections, Rwanda; ▥, Pregnant women, Kinshasa, Zaire. *b*, Longitudinal trends in cases of AIDS reported to WHO in Brazil, Kenya and Rwanda.

AIDS (effectively at birth). The per capita rate at which adults acquire infection, λ , is assumed to be given by $\lambda = \beta c Y/N$ where β is the probability of acquiring infection from any one infected partner, c is the average rate of acquiring partners (c is not the mean number, but is the mean, m , plus the variance to mean ratio, σ^2/m , of the relevant distribution of partner-change rates^{19,26}), and Y/N is the probability that any one partner is infected. More generally for heterosexual transmission we should deal with two distinct populations, N_1 of males and N_2 of females, with females acquiring infection from males at a rate $\beta_1 c_1$, and vice versa for female-to-male transmission at a rate $\beta_2 c_2$. However, for simplicity we assume that $\beta_1 c_1 \approx \beta_2 c_2 \approx \beta c$ such that the two-sex model collapses to equations (1) and (2). This appears to be a reasonable assumption as we have shown elsewhere¹⁹ that in the early stages of the epidemic the ratio of the number of seropositive males to seropositive females is roughly $(\beta_1 c_1 / \beta_2 c_2)^{1/2}$. Observations in a variety of African countries suggest these rates are roughly equal^{3,19}.

To clarify discussion we introduce the following notation: r is defined as the growth rate of the population before the initial HIV infection, Λ as the initial exponential growth rate of the infection within the population, and θ as the extra mortality linked to infection (arising from horizontal and vertical transmission). These rates are given by

$$r = \nu - \mu, \Lambda = \beta c - (\mu + \alpha), \theta = \alpha + \nu(1 - \varepsilon) \quad (4)$$

The model defined by equations (1) and (2) has an exact solution for $N(t)$ and for the prevalence of infection $y(t) (= Y(t)/N(t))$ where

$$N(t) = N(0)e^{rt}[1 + (b/a)\Delta(e^{at} - 1)]^{-\theta/b} \quad \text{and} \quad (5)$$

$$y(t) = [\Delta \exp(at)] / [1 + (b/a)\Delta[\exp(at) - 1]] \quad (6)$$

Here $N(0)$ and $Y(0)$ (where $\Delta = Y(0)/N(0)$) are respectively the initial total population and number of infecteds at time $t = 0$. The parameter combinations a and b are defined as

$$a = \Lambda - r \quad \text{and} \quad b = a + \varepsilon \nu \quad (7)$$

The model exhibits three patterns of behaviour. The basic reproductive rate of infection, R_0 , defined as the number of secondary cases of infection typically generated by one primary case in a susceptible population, is given by (see refs 19, 26):

$$R_0 = \beta c / (\mu + \alpha) \quad (8)$$

Hence if $R_0 < 1$ (which implies $\Lambda < 0$, see equation (4)) the infection cannot establish and there will be no epidemic (note that equation (8) defines the control problem: sexual behaviour, c , must be changed such that $R_0 < 1$). Given the rapid spread of HIV in certain countries, of more importance are the two cases that arise when $R_0 > 1$. If $\Lambda/\beta c > r/\theta$, mortality associated with infection is so high that deaths eventually exceed births and the population begins to decline. In principle, therefore, sexually transmitted infections that also spread via vertical transmission are capable of causing the extinction of their host population.

Alternatively if $0 < \Lambda/\beta c < r/\theta$ the population will eventually settle to exponential growth, the infection being maintained within it, provided $R_0 > 1$. The critical combination of parameter values that divide the two cases is when

$$\varepsilon \nu = (\mu + \alpha)(\Lambda - r) / \Lambda \quad (9)$$

When the right-hand side exceeds the left-hand side, extinction occurs; in the opposite case, sustained growth occurs. Note that extinction cannot occur unless $\Lambda > r$; that is, the exponential rate at which the infection initially spreads must exceed the overall population growth rate, otherwise a decreasing fraction will experience infection, even though the absolute number of infected individuals is growing. When the population continues to grow with the infection being maintained, the prevalence of infection approaches a constant value, y^* , (as $t \rightarrow \infty$) where

$$y^* = (\Lambda - r) / [(\Lambda - r) + \varepsilon \nu] \quad (10)$$

The asymptotic exponential rate at which $N(t)$ and $Y(t)$ grow, ρ (provided the fraction infected who develop AIDS, f , is equal to 1), is

$$\rho = -(\mu + \alpha) + [\varepsilon \nu(\Lambda + \mu + \alpha)] / [\Lambda - r + \varepsilon \nu] \quad (11)$$

In the case where $\rho < 0$ and population growth is eventually halted by AIDS, the period of time before the population ceases its previously exponential growth, t_c , is given by:

$$t_c = \left(\frac{1}{\Lambda - r} \right) \ln \left(\frac{r[1 - \Delta(b/a)]}{\Delta[\theta - r(b/a)]} \right) \quad (12)$$

where Δ is the fraction infected at $t = 0$.

Time-delayed recruitment

The basic model dealt with the total population. In reality, however, HIV spreads predominantly among sexually active adults, and births occur only from sexually mature females, so that it is more realistic to interpret $N(t)$ as the population of sexually active adults (subdivided into $X(t)$ and $Y(t)$). If τ is the average time taken to attain sexual maturity (say 15 years), the relationship between the average birth rate (ω) expressed per adult, and the average birth rate per member of the total population (ν) is:

$$\nu = \omega e^{-(\mu + r)\tau} \quad (13)$$

Table 1 AIDS cases reported to WHO as of September 1987 and demographic parameters for a selection of countries

	Kenya	Zaire	Uganda	United Kingdom
Total number of AIDS cases	625	335	1,138	935
Rate per 1,000 population	0.028	0.010	0.071	0.017
Time period to which demographic data apply	1984–85	1984–85	1984–85	1984–85
Crude population size 1987 (millions)	22.397	31.796	16.018	55.678
Crude live birth rate per 1,000 population yr ⁻¹	55.1	44.8	50.3	12.9
Crude per capita birth rate yr ⁻¹ , ν	0.0540	0.0442	0.0495	0.0129
Life expectancy at birth (yrs)	52.9	52.0	51.0	73.7
Crude per capita death rate yr ⁻¹ , μ	0.0189	0.0192	0.0196	0.0136
Population growth rate per 1,000 yr ⁻¹	41.1	30.3	33.5	1.5
Crude per capita growth rate yr ⁻¹ , r	0.0403	0.0299	0.033	0.0015
Urban population (% of total)	15.5	36.6	9.5	87.7
Dependency ratio (ages < 15, > 64)/(ages 15–64)	1.190	0.928	1.023	0.529
Child dependency age < 15/(ages 15–64)	1.150	0.871	0.973	0.298
Population density, km ⁻²	38.0	13.6	68.0	230.0

From World Bank and United Nations data series.

Here $e^{-\mu\tau}$ defines the proportion of new births that survive to sexual maturity at age τ with mortality rate μ ; the factor $e^{-r\tau}$ measures the sexually mature fraction of the population (assuming type II survival). The demography and epidemiology of HIV in the adult population is now described by equations (1), (2) and (13) but with the original ν replaced by the ν defined in equation (13). The time delays make these equations significantly more complicated and no exact solution is possible. However, we get the earlier results for ρ (equation 11) and for y^* , the asymptotic fraction infected (equation 10) with ν now being replaced by:

$$\nu \rightarrow \omega e^{-\mu\tau - \rho\tau} \quad (14)$$

The replacement is also made in $r = \nu - \mu$. This again leads to equation (9) being the critical condition dividing population growth ($\rho > 0$) from population decline ($\rho < 0$), with the proviso that ν in equation (9) is interpreted as $\omega e^{-\mu\tau}$ and r is interpreted as $(\omega e^{-\mu\tau} - \mu)$.

Epidemiological parameters

The models reveal that whether or not AIDS will change positive population growth rates to negative ones, and the timescale of such changes, depends critically on the magnitudes of various key demographic and epidemiological parameters. The demographic parameters (birth, death and population growth rates, ν , μ , and r respectively) for a series of countries in which HIV is spreading rapidly (Kenya, Zaire and Uganda) are recorded in Table 1. For comparison, data from the United Kingdom are also presented. Note the very high population growth rates in the developing countries (4% per annum in Kenya). As mentioned earlier, epidemiological data are extremely limited and in certain areas we can only guess at parameter ranges. The fraction of those infected who will eventually develop AIDS, f , appears likely to be very high as long-term studies of infected homosexual men in developed countries reveal no decrease in the rate (2–7% per annum) of conversion from HIV infection to AIDS^{38–41}. Disease progression in African heterosexuals appears to occur at similar rates^{11,42}. Roughly 30–40% of infected male homosexuals appear to develop AIDS on an 8–9 year timescale^{5,39,41}, and it seems likely that most of the rest will convert to disease over a longer period (perhaps 15–20 years). Limited studies from North America and Europe have recorded HIV infection in 30–65% of babies of HIV seropositive mothers^{12–14,43–45} and similar studies in Nairobi, Kenya and Kinshasa, Zaire, suggest figures of 46–51%¹¹. *In utero* infection appears to be the dominant mode of vertical transmission and death rates among infected babies are very high—nearly 20 times that of children born to uninfected mothers in one study^{2,11}. Data on the transmission probability, β (either from men to women or vice versa), in heterosexuals and the effective

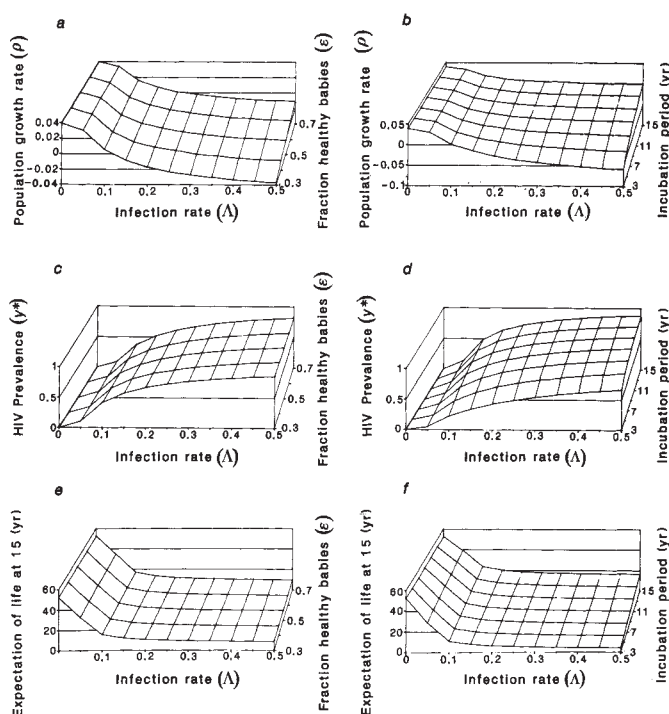


Fig. 2 *a* and *b*, The asymptotic population growth rate per head, (ρ) *c* and *d*, the asymptotic prevalence, that is the fraction of the population infected (y^*); *e* and *f*, the expectation of life at age 15 years. Each graph refers to a human population infected with HIV-1 with varying values for the epidemiological parameters λ (rate of infection per capita per year), ϵ (fraction of babies born to infected mothers who do not acquire infection via vertical transmission) and $1/(\alpha + \mu)$ (average incubation (= infectious) period). Predictions are based on the simple delayed-recruitment model described in the main text. In graphs *a*, *c* and *e* the duration of the incubation (= infectious) period was set at eight years. In graphs *b*, *d* and *f* the fraction of healthy babies born to infected mothers was set at 0.5.

rate of partner change, c , are very limited at present for developing countries. One study suggests a value of β in the range 0.05–0.1 per partner for female to male transmission, provided cofactors (genital ulcers) are present⁴⁶. A rough guide to the parameter combination βc can be obtained indirectly from a knowledge of the rate of increase of the infected population^{18,26} (or from the doubling time of the epidemic). The limited data on trends in the rise in seropositivity in specific African countries are summarized in Table 2. The rate of spread is highest in urban areas in Africa^{2,3,11}, and is particularly fast in people with

Fig. 3 *a*, Population trajectories through time (recorded as $\ln(N(t)/N(0))$ where $N(0)$ is the population size at time $t=0$ when the infection was introduced) as predicted by the delayed-recruitment model defined in the main text for various values of the fraction of babies born to infected mothers who do not acquire HIV-1 infection via vertical transmission (ϵ). The top line denotes 4% growth in an uninfected population. The remaining trajectories, from bottom to top, denote predictions with ϵ set at 0.3, 0.5 and 0.7 respectively. The death rate of uninfecteds was set at $\mu = 0.019$ per capita per yr (life expectancy of 52 yrs), incubation period of the disease was set at 8 yrs ($= 1/(\alpha + \mu)$), and the time period to recruitment to the reproductively mature age-class (τ) was fixed at 15 years. The rate of infection Λ was set at 0.233 per capita per yr. *b*, Population trajectories through time, as in (*a*) but varying the fraction of those adults infected, f , who proceed to develop AIDS and die. The top line denotes 4% population growth in an uninfected population. The remaining trajectories, from top to bottom, denote predictions with f set at 0.25, 0.5, 0.75, 0.9, 0.95 and 1 respectively. The inclusion of the fraction f into the delayed-recruitment model defined in the main text is as described in refs 26, 37. Parameter values as in (*a*), with ϵ fixed at 0.5, assuming that an equal fraction of children born to seropositive-recovered women die effectively at birth. *c*, Growth rate r in the uninfected population which can asymptotically be brought to zero (stationary population) for a specified value of f . The other epidemiological and demographic parameters are: $\Lambda = 0.4$ per capita per year; $1/(\alpha + \mu) = 10$ yrs; $\mu = 0.02$ yr $^{-1}$; $\tau = 15$ yrs. The top line is for $\epsilon = 0$ and the bottom for $\epsilon = 0.5$. *d*, The period of time, t_c , predicted before an infected population ceases its previous pattern of exponential growth and begins to decline after invasion by HIV-1 (equation 15) as a function of the rate of spread of infection Λ and the population growth, r .

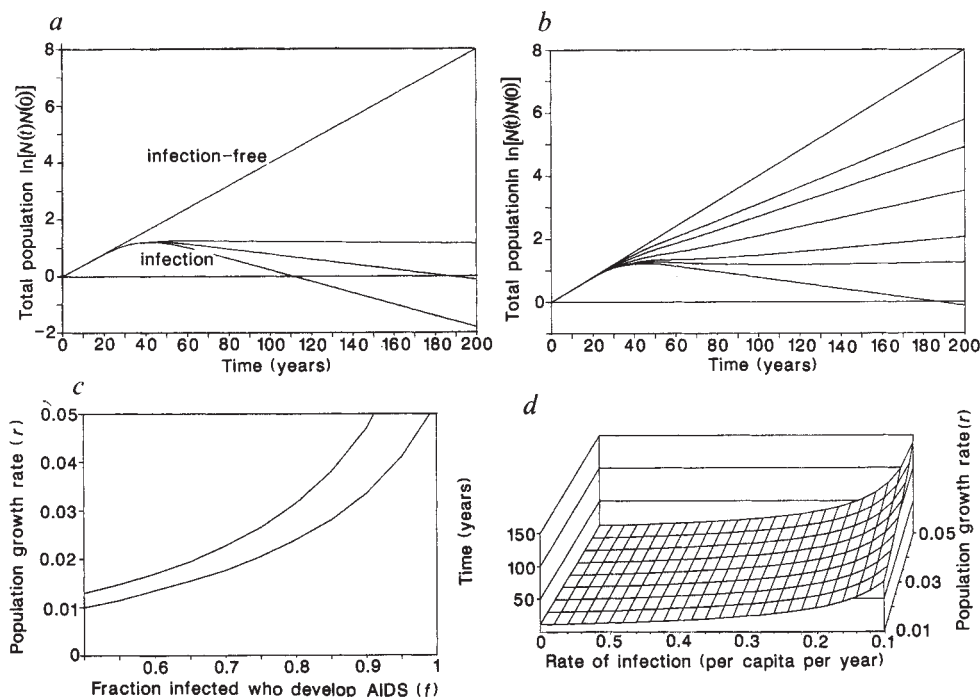


Table 2 Estimates of the rate of spread of HIV infection in various populations in developing countries (see text for details)

Location	Years	Population surveyed	Doubling time t_d (yrs)	Rate of spread of infection, Λ (yr $^{-1}$)	Parameter combination $\beta c(1/(\alpha + \mu)) = 15$ yrs	Basic reproductive rate, R_0 ($1/(\alpha + \mu) = 15$ yrs)	Data source (reference)
Kenya, Nairobi	1981-85	Prostitutes	1.00	0.681	0.748	11.2	53
Kenya, Nairobi	1982-85	Men with chancroid	1.00	0.677	0.744	11.1	53
Kenya, Nairobi	1981-85	Men with STDs*	1.67	0.414	0.480	7.2	53
Kenya, Nairobi	1981-82	Women with STDs	1.00	0.686	0.753	11.3	53
Kenya, Nairobi	1970-86	Pregnant women	2.87	0.241	0.308	4.6	11
Zaire, Kinshasa	1970-86	Mothers	3.50	0.198	0.265	4.0	54
Uganda, Kampala	1985-87	Women—antenatal	1.12	0.619	0.686	10.3	55

* STD, sexually transmitted disease.

multiple sexual partners, in those who have a high rate of infection with other sexually transmitted diseases (STDs) such as chancroid¹⁶, and in prostitutes^{2,3,47,48}. For example, the doubling time (t_d) of the epidemic among prostitutes in Nairobi^{2,48} from 1981 to 1985 was approximately 1.0 years. In the general urban populations in Kenya, Uganda and Zaire (as indicated by the proportion of pregnant women attending health clinics, found to be infected^{2,3,11}) the rate of spread is slower, but still alarming, doubling times ranging from 1 to 3.5 years (Table 2). Assuming equality of transmission from female to male and vice versa, given the observed 1:1 sex ratio in seropositivity^{2,3,11} and an average incubation period (= average infectious period) of roughly 15 years ($1/(\alpha + \mu)$) these doubling times in the general population yield estimates of R_0 (the number of people infected by a single seropositive individual) and βc (the rate of transmission) in the respective ranges of 4–12 and 0.2–0.8 (refs 17,26,37). The associated estimates of the rate of increase (Λ) in the infected population ($Y(t)$), in the early stages of the

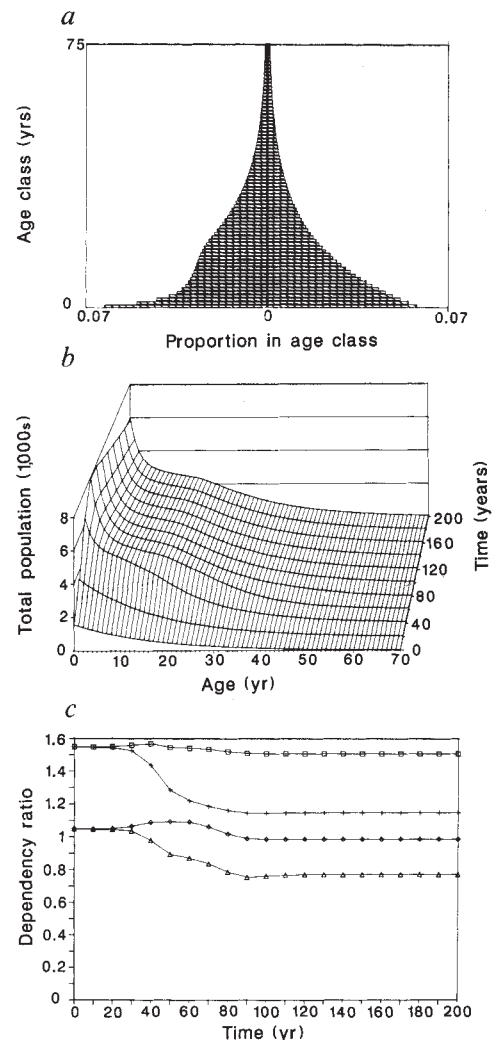
epidemic in different countries/populations are recorded in Table 2 (range 0.2–0.8 yr $^{-1}$).

On the basis of these estimates we choose the following parameter ranges to explore the properties of the recruitment-delay version of the model: $\epsilon = 0.3$ –0.7; $1/(\alpha + \mu) = 8$ –20 yr; $\mu = 0.019$ yr $^{-1}$; $\Lambda = 0.2$ –0.5 yr $^{-1}$ (for the general urban population); $\tau = 15$ yrs; $\nu = 0.02$ –0.06 yr $^{-1}$ (where ν is the birth rate per member of the population, before AIDS). Note that from Tables 1 and 2 it appears that Λ always exceeds r in value; hence the infection is predicted to have a severe effect on population growth.

Population growth rates

A wide range of parameter values, all within the bounds suggested by current empirical studies, predicted asymptotically negative population growth rates even when recruitment delays are taken into account. The relationship between the asymptotic growth rate, a range of values for the rate of infection, Λ , and

Fig. 4 The predictions presented in graphs *a* to *c* are derived from a fully age-structured model with the following partial differential equations for the numbers susceptible ($X(a, t)$), numbers infected ($Y(a, t)$) and the total population ($N(a, t)$) of age a at time t : $\partial[X(a, t)] = -[\lambda(a, t) + \mu(a)]X$; $\partial[Y(a, t)] = \lambda(a, t)X - [\alpha(a) + \mu(a)]Y$; $\partial[N(a, t)] = -\mu(a)N - \alpha(a)Y$ where the operator $\partial[F]$ is short-hand for $\partial F/\partial t + \partial F/\partial a$. One boundary condition is the requirement $X(0, t) = \int_0^\infty m(a)[N(a, t) - (1 - \varepsilon) \times Y(a, t)]da$, and $Y(0, t) = \int_0^\infty m(a)(1 - \varepsilon)Y(a, t)da$ with $m(a)$ defining the age-specific fertility rate. The other boundary condition is given by specifying $X(a, 0)$, $Y(a, 0)$ and $N(a, 0)$ at $t = 0$. The force of infection $\lambda(a, t)$ is given by $\lambda(a, t) = \beta c_\tau \int_0^\infty j(a, a')Y(a', t)da' / \int_0^\infty j(a, a')N(a', t)da'$ where β and c are the transmission probability and the mean rate of acquiring new sexual partners and $j(a, a')$ defines the probability that a susceptible of age a will choose a partner of age a' . For illustration we make the simplifying assumption that the death rate $\mu(a) = \mu$ (reasonable for developing countries), the birth rate $m(a) = \omega$ for $T > a > \tau$, $m(a) = 0$ otherwise (where τ defines the lower limit for reproduction and T the upper limit, set at 15 and 50 years respectively) and $j(a, a')$ has a constant value (that is, sexually active adults choose partners independent of age). We assume that the fraction of children born to infected mothers that develop AIDS ($1 - \varepsilon$) survive for an average of two years from birth. In these circumstances we obtain equation (11) in the main text with ν defined as $\omega \exp[-(\mu + \rho)\tau]$, where ω is the birth rate per adult, for the asymptotic growth rate ρ of the population. The asymptotic prevalence $Y/N = (\Lambda - \rho)/\beta c$ where Λ is the initial growth rate of the seroprevalence of HIV. The fraction of infecteds who develop AIDS, f , was assumed to be 1. The model allows us to calculate time-dependent changes, and asymptotic results for the age profile of infected and uninfected populations. *a*, Asymptotic age distributions (fraction in each age class) of an uninfected (right-hand side) and an infected (left-hand side) population. The growth rate (r) is 4% and life expectancy ($1/\mu$) is 52 years in the absence of HIV. The incubation period was set at 15 years ($1/(\alpha + \mu)$) with the fraction of healthy babies born to infected mothers (ε) set at 0.3. The rate of infection in the early stages, Λ , was set at 0.233 yr^{-1} . *b*, Temporal changes in the age distribution of a population as the growth rate of the population changes from positive to negative as HIV spreads (parameter values as for left-hand side of *a*). *c*, Temporal changes in the dependency ratio ((ages < 15 and > 64)/ages 15–64) as HIV spreads. Four different numerical simulations of age-structured models are recorded. In all, $\tau = 15 \text{ yr}$, $\mu = 0.019$ per capita per yr, $1/(\alpha + \mu) = 15$ years; parameters (r) and (ε) are varied as follows: ($\square$) $r = 0.04$ per capita per yr, $\varepsilon = 0.7$; (+) $r = 0.04$ per capita per yr, $\varepsilon = 0.3$; ($\diamond$) $r = 0.02$ per capita per yr, $\varepsilon = 0.7$; ($\triangle$) $r = 0.02$ per capita per yr, $\varepsilon = 0.3$.



the fraction of babies born to infected mothers who do not acquire HIV via vertical transmission, ε , is recorded in Fig. 2 for various values of the incubation period. In these examples the growth rate per head of population before HIV infection, r , was assumed to be 0.04 yr^{-1} (4% growth per annum) with life expectancy 52 years. As ε decreases and Λ increases the asymptotic growth rate is predicted to become negative. Concomitantly, as $\varepsilon \rightarrow 0$ and Λ rises, the asymptotic prevalence of HIV infection is predicted to rise in the adult population (>15 yrs) to high levels (Fig. 2). For plausible ranges of parameter values, HIV infection resulting in AIDS is predicted to induce the severe changes in the population size shown in Fig. 3*a* and *b*. The growth rate in population before any HIV infection, r , which can asymptotically be brought exactly to zero (stationary population) for a specified value of the fraction infected who develop AIDS, f , is shown in Fig. 3*c*.

The period t_c , denoting the time taken before the population begins to decline after invasion by HIV, is given very approximately by

$$t_c \sim [\ln(1/\Delta)]/[\Lambda - r] \quad (15)$$

where Δ is the fraction infected at time $t = 0$ (ref. 37). This relationship is depicted in Fig. 3*d* for a range of values for the pre-HIV infection population growth rate, r (chosen to mimic developing countries—see Table 1), and the rate of HIV infection Λ (see Table 2). Note that for low to moderate infection rates (like those observed in the general population in Zaire, Uganda and Kenya whose $\Lambda = 0.1\text{--}0.2 \text{ yr}^{-1}$) the time to the onset of population decline, t_c , is predicted to be very long, (20–70 yrs)

even in countries with low growth rates ($r = 0.02 \text{ yr}^{-1}$). Following onset, asymptotic rates of population decline may be dramatic (Fig. 2).

Changes in age structure

The simple model defined above can easily be extended to incorporate a variety of realistic refinements. These include: asymmetric probabilities of transmission from females to males, and males to females; asymmetries in the probability that a male or female partner will be infected, reflecting age dependency in the choice of partners of the opposite sex; only a fraction of those infected proceeding to develop AIDS; and, most importantly, a full age structure to reflect the demographic impact of AIDS on the age-distribution of infected populations^{17,26,37}. The details of an age-structured model are defined in the legend to Fig. 4 in which predicted changes in the asymptotic age distribution of an infected population for a specific set of demographic and epidemiological parameters are recorded. The age profiles of the population before and after the establishment of HIV infection are represented as the relative proportions in different age classes (as opposed to absolute numbers in the declining infected population) (Fig. 4*a*). The effect of AIDS upon the dependency ratio—which is taken to be the population below age 15 and above age 64 years, divided by that aged 15 to 64 years—is not as great as might be expected, or has been widely suggested^{2,3,11}. On the one hand, the direct effects of mortality in the sexually active adult age classes due to AIDS tends to increase the ratio. On the other hand, the general depression of overall population growth rates due to adult deaths, and to the

reduction in effective birth rates due to the deaths of infected babies, tends to decrease the ratio. The net outcome depends on the precise values of the demographic and epidemiological parameters. An important prediction is that, for some plausible ranges of values, AIDS would have little effect on, or even slightly decrease, the dependency ratio (Fig. 4c), a result in sharp contrast with the views of many informed individuals who have discussed this problem^{2,3,11,49}.

Discussion

Our analyses, based on very simple models of the major demographic and epidemiological processes, yield three major conclusions. First, for plausible ranges of parameter values the disease AIDS is predicted to be capable of significantly reducing population growth rates, and even depressing them to negative values. Second, the time for these effects to become fully manifest after the initial introduction of HIV infection, is predicted to be long, of the order of many decades. Third, whether or not AIDS will decrease or increase the dependency ratio within a given population depends critically on the values of the major demographic and epidemiological parameters prevailing in the community. For plausible parameter ranges (Tables 1 and 2) our analyses suggest little change or a small, but beneficial, change. This conclusion appears fairly insensitive to variations in the fraction of those infected (f) who eventually die. Similar analyses of the impact of directly transmitted infections such as smallpox and bubonic plague, that were of great historical significance as causes of human morbidity and mortality, suggest that AIDS has greater potential to depress significantly human population growth rates^{28,29}. This stems from the ability of HIV to transmit both horizontally via sexual contact and vertically from mother to unborn offspring, the high mortality associated with infection and the apparently long period over which infected persons are asymptomatic but infectious to their sexual partners.

Before proceeding to discuss these conclusions further, we emphasize that they derive from deliberately simplified models and from limited empirical data. As data accumulate, future numerical studies of more complex and realistic models must address a series of key problems. Of major importance is the assumption of homogeneous mixing made when estimating the rate of spread of infection by sexual contact. Recent theoretical and empirical studies challenge this assumption, recording marked heterogeneity, dependent on age, sex and socio-economic factors, in the frequency distributions of rates of sexual partner change^{18,26,27}. Equally important is the question of promiscuity class (namely, do individuals in high sexual-partner change classes predominantly choose partners in their own activity class?), and the role of prostitutes in the spread of infection within the general population³⁷. Simple cascade models in which infection is assumed to flow from a core group of highly sexually-active female prostitutes, to males (assumed all to be moderately promiscuous) and thence to the child-bearing women (assumed all to be relatively monogamous) reveal properties qualitatively similar to those outlined earlier³⁷; the quantitative details depend on the actual transmission probabilities and rates of partner change among these three groups in, say, Africa, which are presently uncertain. Other problems concern infectiousness throughout incubation of the disease^{17,18,24}, the role of cofactors in transmission, and the assumption of equal transmissibility between the sexes. The overall average transmission rate, $\bar{\beta}c$, is the geometric mean of the average transmission rates β_1c_1 (male to female) and β_2c_2 (female to male), and $\bar{\beta}c = (\beta_1c_1\beta_2c_2)^{1/2}$ (refs 19,37). The symmetric assumption of our simple model ($\beta_1c_1 \approx \beta_2c_2$) is a reasonable approximation of the more general asymmetric case ($\beta_1c_1 \neq \beta_2c_2$) provided $\beta c \gg (\mu + \alpha + r)$; this appears to be the case for the representative parameter values considered (Table 2, Fig. 2).

In general, heterogeneity in sexual activity, whether between urban and rural areas, or between different groups in a given

community, is likely to decrease the demographic impact predicted by our simple model. For any chosen values of the early doubling rate and incubation interval, the eventual prevalence of endemic infection will be higher if all individuals are equally active sexually than if much variability exists (in which case core infection tends to be concentrated within the more active categories)^{36,50}. Our parameter values, based on current data, may well lead to conclusions more representative of, say, Kinshasa and Nairobi than of Zaire and Kenya as a whole. But any assessment of the effects of such heterogeneities simply requires better data than we have at present.

Simple models, such as those described here, are useful for setting the agenda for data collection to improve the accuracy of predictions. There is a clear need for the collection of much more detailed data on all the parameters discussed above, although the social, ethical and practical problems surrounding such research are formidable. A start has been made with recent theoretical⁵¹ and empirical⁵² work that has begun to address some of the important complications induced by heterogeneity in rates of sexual partner change, and by complex networks of sexual and social interactions.

With these limitations in mind, we conclude by returning to two major predictions of the simple model. First, that the AIDS epidemic is likely to have little impact on dependency ratios, a prediction which directly contradicts current views of this problem^{2,11}. From this it might be concluded that the epidemic will be less disruptive to the social organization and economic fitness of badly afflicted countries than previously feared. We do not believe this to be the case, however, as the very high predicted mortality due to a disease which requires repeated hospitalization, perhaps over periods of a few years, and which is thought to enhance morbidity due to other infections such as tuberculosis², will be devastating to already overloaded health-care systems in poor countries^{2,11}. Small changes in the dependency ratio (whether plus or minus) are irrelevant under such circumstances.

The second prediction, that of a long time-period between the establishment of HIV-1 infection and the reversal of exponential population growth, offers hope in the sense that it provides time for educational programmes to change behaviour¹¹, for the development of more effective drugs and (we hope) for the discovery of an effective vaccine. It should be remembered, however, that the availability of cheap and effective vaccines (such as that against measles) is not always beneficial to people living in very poor countries where the cost of immunization is prohibitive to individuals and governments. For the foreseeable future the main hope for checking the spread of HIV lies in the development and forceful application of education programmes aimed at changing behaviour. The predictions of our simple models highlight the urgency of implementing such programmes.

We thank the Overseas Development Agency and Panos for financial support (R.M.A. and A.R.M.) and the NSF (R.M.M.). We have greatly benefited from conversations with A. Johnson, P. Piot, F. Plummer, J. Bull, G. Medley, S. Blythe, E. Konings, M. John and J. Bongaarts.

Received 11 December 1987; accepted 22 February 1988.

1. World Health Organization (WHO). *Wkly Epidemiol. Rec.* **62**, 301 (1987).
2. Piot, P. & Carael, M. *Br. med. Bull.* (in the press).
3. Quinn, T. C. *et al. Science* **234**, 955-965 (1986).
4. Brun-Vezinet, F. *et al. Science* **234**, 955-963 (1984).
5. Mann, J. M. *et al. Lancet* **ii**, 707-709 (1986).
6. Wendler, I. *et al. Br. med. J.* **293**, 782-785 (1986).
7. Biggar, R. J. *Lancet* **i**, 79-82 (1986).
8. Melbye, M. *et al. Lancet* **ii**, 1,114-1,115 (1986).
9. Piot, P. *et al. J. infect. Dis.* **155**, 1,108-1,112 (1987).
10. Mann, J. M. *et al. Lancet* **ii**, 654-656 (1986).
11. Piot, P. *et al. Science* **239**, 573-579 (1988).
12. Coutinho, R. A. *et al. Ned. Tijdschr. Geneesk* **130**, 508 (1986).
13. Melbye, M. *et al. Ann. intern. Med.* **104**, 496-451 (YEAR?).
14. Weber, J. N. *et al. Lancet* **i**, 1,179-1,181 (1986).
15. Kanki, P. J. *AIDS* **1**, 141-145 (1987).
16. Quinn, T. C. *et al. J. Am. med. Assoc.* **257**, 2,617-2,621 (1987).
17. May, R. M., Anderson, R. M. & Johnson, A. M. *Science* (submitted).

18. Anderson, R. M. *J. R. statist. Soc. A* (in the press).
19. May, R. M. & Anderson, R. M. *Nature* **326**, 137-142 (1987).
20. Piot, P. & Mann, J. M. *Annls Inst. Pasteur, Paris* **138**, 125-132 (1987).
21. Mann, J. *et al. New Engl. J. Med.* **316**, 345 (1987).
22. Kreiss, J. *et al. New Engl. J. Med.* **314**, 414-418 (1986).
23. D'Costa, I. J. *et al. Sex. transm. Dis.* **12**, 64-67 (1985).
24. Medley, G. F. *et al. Nature* **328**, 719-721 (1987).
25. Mann, J. M. *et al. J. Am. med. Assoc.* **255**, 3,255-3,259 (1986).
26. Anderson, R. M. *et al. IMA. J. math. med. Biol.* **3**, 229-263 (1986).
27. Anderson, R. M. *et al. Lancet* **i**, 1,073-1,075 (1987).
28. Anderson, R. M. & May, R. M. *Nature* **280**, 361-367 (1979).
29. May, R. M. & Anderson, R. M. *Nature* **280**, 455-461 (1979).
30. Anderson, R. M. in *Theoretical Ecology: Principles and Applications* (ed. May, R. M.) 318-355 (Blackwell, Oxford, 1981).
31. Anderson, R. M. *Lect. Math. Life Sci.* **12**, 1-68 (1979).
32. May, R. M. & Anderson, R. M. *Math. Biosci.* **76**, 1-16 (1985).
33. McLean, A. R. & Anderson, R. M. *Epidem. Infect.* **100**, 111-133 (1988).
34. Anderson, R. M. & May, R. M. *Phil. Trans. R. Soc. B* **314**, 533-570 (1986).
35. Isham, V. *J. R. statist. Soc. A* (in the press).
36. Hethcote, H. W. & Yorke, J. A. *Lect. Notes Biomaths.* **56**, 1-105 (1984).
37. May, R. M., Anderson, R. M. & McLean, R. M. *Math. Biosci.* (in the press); *Lect. Notes Biomaths.* (in the press).
38. Brodt, H. R. *et al. Dt. med. Wschr.* **111**, 1,175-1,180 (1986).
39. Kaplan, J. E. *et al. J. Am. med. Assoc.* **257**, 335-339 (1987).
40. Goedert, J. J. *et al. Lancet* **ii**, 711-715 (1984).
41. Zolla-Pazner, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5,404-5,408 (1987).
42. Plummer, F. A. *et al. Lancet* **i**, 1,293-1,295 (1983).
43. Saulsbury, F. T. *et al. Pediat. infect. Dis. J.* **6**, 544-549 (1987).
44. Borkowsky, W. *et al. Lancet* **i**, 1,168-1,171 (1987).
45. Mok, J. Q. *et al. Lancet* **i**, 1,164-1,168 (1987).
46. Cameron, D. W. *et al. Abstr. 3rd int. Conf. AIDS* (Washington DC, 1987).
47. Van De Pierre, P. *et al. Lancet* **ii**, 524-527 (1985).
48. Piot, P. *et al. J. infect. Dis.* **155**, 1,108-1,112 (1987).
49. Mann, J. J. *New Scient.* **113**, 40-43 (1987).
50. May, R. M. *Ecology* **67**, 1,115-1,126 (1986).
51. Dietz, K. & Hader, K. P. *J. Math. Biol.* **26**, 1-25 (1988).
52. Klov Dahl, A. S. *Soc. Sci. Med.* **21**, 1,203-1,216 (1985).
53. Brun-Vezinet *et al. Science* **226**, 453-456 (1984).
54. John Hopkins University, *Population Reports 7 Series L* No. 6, 10-14 (1987).
55. Wilson Carswell, J. & Lloyd, G. *AIDS* **1**, 192-193 (1987).

LETTERS TO NATURE

Gravitational radiation and gravitational lensing as a source of electromagnetic bursts

B. McBreen* & L. Metcalfe†

* Physics Department, University College Dublin, Dublin 4, Ireland

† Astrophysics Division, Space Science Department of ESTEC, The Netherlands

What causes gamma-ray bursts has long been a puzzle to theoreticians. Here we propose that fluctuations in the position of a compact source originating in a stochastic background of gravitational waves cause the source to pass back and forth across the high-intensification caustics of an intervening galactic lens. The intensified images cross the Solar System with superluminal speed and give rise to gamma-ray bursts (GRBs)¹⁻³. The features of this model that distinguish it from other models⁴ of GRBs are (1) the GRB is a result of a lensing phenomenon and not a sudden power surge in a compact source; (2) the directions to the sources obtained with the light-travel-time (LTT) method are incorrect; (3) the spectral features arise from high-angular-resolution scans across accretion disks around Kerr black holes; and (4) the lensed sources are probably BL Lac objects.

A stochastic background of long-wavelength gravitational waves can be detected by small fluctuations in the position or jitter of a light beam reaching an observer if the wavelength of the gravitational waves is much larger than the source⁵. The random fluctuations in the position of a source are given by⁵

$$\langle \theta^2 \rangle^{1/2} \approx 0.2 \frac{\lambda_G}{\lambda_H} \left(\frac{\rho_G}{\rho_C} \right)^{1/2} \quad (1)$$

where λ_G is the wavelength of the quasi-monochromatic gravitational waves, $\lambda_H \approx 10^{28}$ cm (the Hubble radius), ρ_G the energy density in gravitational waves and ρ_C the closure density of the universe. $\langle \theta^2 \rangle^{1/2}$ is independent of the distance to the source because of the transverse nature of gravitational radiation. The timescale for a complete reversal of the source position is of the order of the wave period. The magnitude of $\langle \theta^2 \rangle^{1/2}$ is 4×10^{-6} arc s at $\lambda_G = 10^{18}$ cm and $\rho_G = \rho_C$. Gravitational waves on the Mpc scale will cause jitter in the source position⁶ of magnitude 0.1 arc s for $\rho_G = 10^{-4} \rho_C$. The distribution of amplitudes and frequencies in the gravitational wave background is largely unknown as only upper limits have been established⁷⁻⁹ at a few widely different values of λ_G . Several mechanisms contribute to an isotropic cosmological background of gravitational waves that may be an appreciable fraction of the total energy content

of the Universe^{10,11}. The only confirmed source of gravitational waves¹² is binary stars at a level of $\sim 10^{-7} \rho_C$.

The collision interaction of gravitational waves is a fundamental problem in gravitational theory and is important because gravitational waves generated at redshift $z > 1$ interact. Exact solutions describing colliding gravitational waves exhibit space-time singularities due to the mutual focusing of the two waves¹³. It is possible that such amplification is occurring and producing apparent fluxes and values of $\langle \theta^2 \rangle^{1/2}$ larger than that given by equation (1). The interaction of gravitational waves may have been important in the early Universe and created a space-time network of caustics. This network could have given rise to the density fluctuations necessary for galaxy formation. Note that the distribution of galaxies in redshift space appear³ to follow a space-time network¹⁴.

The predicted values^{5,6} of $\langle \theta^2 \rangle^{1/2} \approx 10^{-6}$ arc s are too small to be measured by standard astronomical techniques but are within the capability of a galactic gravitational lens¹⁵. If the compact source, lens and Earth are aligned to within 10^{-6} arc s of the caustic, fluctuations in the source position due to the gravitational wave background cause the source to move back and forth across the high-intensification caustic.

The speed of the intensified image at the Earth is

$$\langle v^2 \rangle^{1/2} = \frac{2 \langle \theta^2 \rangle^{1/2} c D}{\lambda_G}$$

where D is the distance to the lens and c is the velocity of light. The image speed, v , will be superluminal, for lenses at cosmological distances, and fluctuations that exceed $\langle \theta^2 \rangle^{1/2}$ by a factor of three and also for smaller fluctuations that occur on a time-scale short compared with the period of the gravitational waves. A shift in the apparent position of the source at the lens by 10^{-6} arc s results in a displacement of the image by $\sim 5 \times 10^{16}$ cm in the vicinity of the Solar System. Superluminal intensified images may also result from sources crossing the caustics of rapidly moving cosmic strings or rapidly moving sources crossing galactic caustics¹⁵.

The frequency of quasars crossing galactic caustics has been evaluated¹⁵ and is too small by a factor of 10^3 to account for the observed value of 100 GRBs per year. A different result is obtained when microlensing by stars in the Galaxy is considered¹⁶⁻¹⁹ because there are about as many microlenses as stars in the Galaxy and the intensification of the source crossing the microcaustic is strongly dependent on the source size. For objects with solar mass the intensification varies from 10^2 to 10^6 for source sizes between 3×10^{15} cm and 3×10^7 cm. If only gamma rays originate from a small source or cluster of small sources the intensification can exceed that in the optical region by a factor of 10^4 . The frequency of a source crossing the microcaustics depends on the optical depth to microlensing^{17,19} and has a typical value¹⁷ of 0.08 yr^{-1} . The number of sources

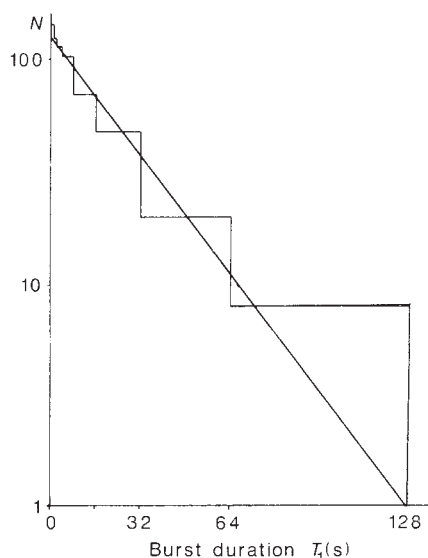


Fig. 1 The distribution of the frequency of bursts, $N(>T_1)$, with duration, T_1 , for 143 events recorded²⁵ by the Venera space probes. The equation of the straight line is given by $N(>T_1) = N_0 e^{-T_1/\tau}$ where $N_0 = 125$ and $\tau = 25$ s. The distribution is consistent with a random process.

required to give 100 GRBs per yr over the whole sky is 1.2×10^3 and is approximately equal to the number of X-ray-selected²⁰ BL Lac objects with fluxes $>10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2}$. It has been proposed²¹ that the unusual properties of BL Lac sources can be explained by microlensing. Furthermore, the optical data display²² rapid changes in the position angle of polarization and are sometimes explained by independent luminosity variations of many highly polarized components. If BL Lac objects at redshifts, ~ 1 , and luminosities of 10^{44} – $10^{46} \text{ erg s}^{-1}$ are intensified by a factor of 10^6 they generate GRBs with peak fluxes, P , in the observed range of 10^{-4} – $10^{-6} \text{ erg s}^{-1} \text{ cm}^{-2}$. This model is different from the macrolensing²³ or microlensing²⁴ of explosive events such as supernova explosions because the microcaustic-crossing part of the Solar System at superluminal speed generates the burst without any inherent change in luminosity of the source.

The lens model is compatible with many of the unusual features observed in GRBs. For example, (1) the directions to GRB sources have been obtained using the light-travel-time (LTT) method and also using the cosine-law response of the onboard detectors^{25,26}. In the lens model the finite value of v (speed of the image) introduced a tilt of the wavefront of magnitude $\phi = \tan^{-1} c/v$ and hence the LTT technique will not give the correct direction to the source. In all other models of GRBs the LTT direction is the true direction to the source. It is significant that the error boxes of the two independent methods do not overlap²⁷ for 17 out of 34 GRBs. In the lens model there will be a variable difference between these two directions and the largest observed inconsistencies of $\sim 20^\circ$ yields $v \approx 2c$ and explains the absence of optical candidates in the LTT directions^{28–30}. The prediction that the LTT directions are incorrect is supported indirectly by the spectacular transient GB790305 where significant differences occur in the power spectra obtained with three space probes². In the lens model these differences can be intrinsic to the source because each space probe detects the GRB at slightly different times due to the finite value of v .

(2) The upper limits to the distances to GRB sources³¹, obtained with the photon-photon absorption process, do not apply because the true source size is determined by v and the photon density in the source is smaller than obtained from the image by a factor equal to the intensification of the lens³². The

absorption process does not impose any cosmological limit on the distance to GRB sources. The detection³³ of gamma-rays with $E > 40 \text{ MeV}$ in some GRBs is expected if the source is an emitter of high-energy gamma-radiation because this radiation will be intensified by the lens. For the same reason the lens model can explain the unusual transient observed³⁴ at $E > 10^{14} \text{ eV}$.

(3) The emission and absorption features^{2,3} observed in 20% of GRB spectra give clues as to the nature of the source crossing the microcaustic. The spectral features of the burst GB790622 have been interpreted³⁵ as an absorption feature at 60 keV and an emission feature at 460 keV but the spectrum is also consistent with two emission features³⁶ at 90 keV and 460 keV. Emission features of this kind arise in the radiation from an accretion disk around a black hole where the flux distribution exhibits a strong asymmetry due to Doppler shifts, gravitational redshifts, gravitational lens effect and the angular momentum parameter, a/m , of the Kerr black hole^{37,38}. An external equatorial observer measures the radiation from the inner edge of the disk as having a maximum blueshift of 0.36 and a corresponding redshift of 2.0 for a hole with $a/m = 0$. The 511-keV positron annihilation line emitted by the semi-relativistic plasma³⁹ is split into two components at 798 keV and 170 keV. The emission features occur at 90 keV and 420 keV for $z = 0.9$ and provide evidence that some of the lensed sources are accretion disks around black holes. The emission feature at 45 keV in GB791006 is consistent with an extreme Kerr black hole because the maximum blueshift increases by a small amount but the redshift has a large value. Significant values of the blueshifts and redshifts occur not only to the radiation emitted from the inner edge of the disk but also at distances from the black hole of up to about one hundred times its radius³⁷. The intensity of the radiation depends on the polar angle of the observer. Some intense GRBs have a very hard spectrum and correspond to an observer near the equatorial plane of an extreme Kerr black hole. The emission features will be separated in time as the accretion disk is scanned near the equatorial plane by the high-angular-resolution microcaustic of the lens and separations of this type are observed in GRB spectra^{3,26,33}. The durations of GRBs will be in the range of 10^{-2} – 10^2 s assuming $v > 2c$ and the radiation emitted from the inner regions of accretion disks with hole masses in the range 10^3 – $10^7 M_\odot$.

(4) No GRB has been detected at optical frequencies⁴⁰ but observations at low X-ray energies show the emission to be less intense and longer in duration⁴¹. The radiation from the inner region of the disk can escape in the gamma-ray region but is attenuated at lower frequencies. The X-ray and optical radiation associated with GRBs will be longer in duration and less intensified by the lens because the radiation originates from a larger region of the disk. Similar conclusions apply to the radio emission except that the gamma-ray and radio caustics will not coincide because of plasma refraction in a galactic lens. The magnitude of the separation of the caustics⁴² is comparable with the fluctuations in the source position and hence the detection of radio emission at different frequencies provide a method of measuring the jitter for individual GRBs.

(5) The relationship between the total flux, S , and frequency, $N(>S)$, of GRBs has been studied^{43,44} in the form of a $\log N - \log S$ distribution. The interpretation of the distribution is complicated by the fact that a strong correlation exists between S and the burst duration, T_1 , such that bursts with higher values of S have longer durations. The correlation is expected in the lens model where a range of values of v generate a similar range in T_1 for sources with the same size. The distribution of 143 burst durations is given in Fig. 1 and is consistent with a random process. The duration distribution of a large number of GRBs agree with the predictions of the lens model. The $\log N - \log P$ distribution is significantly steeper than the euclidean value for $P \leq 10^{-6} \text{ erg s}^{-1} \text{ cm}^{-2}$ and a similar result²⁰ was obtained for X-ray-selected BL Lac objects with fluxes $\leq 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2}$.

The distribution of GRBs is consistent with BL Lac objects that are intensified by a factor of 10^6 .

(6) The 1.2×10^3 lensed sources required to explain GRBs arise from a population of 3×10^4 sources assuming a lensing probability⁴⁵ of 4×10^{-2} . If this population of sources accounts for the gamma-ray background, then the lens model should predict the observed value of 2×10^{-5} of the gamma-ray background in GRBs. The probability⁴⁶ of source intensification by a value greater than A is given by $I(>A) = \pi/A^2$ and the contribution of the high-intensification-microcaustic regions to the gamma-ray background is $\sim 4 \times 10^{-2} \pi A^{-1}$ which agrees with the measured value for $A = 6 \times 10^3$. There is reasonable agreement between the observed and predicted values because the experiments²⁶ triggered on GRBs at a level of ~ 0.4 of the background and obtained the total flux by integrating to a level of ~ 0.1 of the background.

The most important prediction of the lens model is the incorrect directions to GRB sources obtained with the LTT technique. Independent methods are required to determine unambiguously the directions to GRB sources.

Received 13 February; accepted 23 October 1987.

1. Cline, T. L. *AIP Conf. Proc.* **115**, 333-342 (1984).
2. Hurley, K. *AIP Conf. Proc.* **115**, 343-351 (1984).
3. Teegarden, B. J. *AIP Conf. Proc.* **115**, 352-366 (1984).
4. Ruderman, M. *Ann. N.Y. Acad. Sci.* **262**, 164-180 (1975).
5. Zipoy, D. M. *Phys. Rev.* **142**, 825-838 (1966).
6. Dautcourt, G. *Astr. Astrophys.* **38**, 341-344 (1975).
7. Hellings, R. W. & Downs, G. S. *Astrophys. J.* **265**, L39-L42 (1983).
8. Anderson, J. D. & Mashoon, B. *Astrophys. J.* **290**, 445-448 (1985).

9. Drever, R. W. P. in *Gravitational Radiation* (eds Deruelle, N. & Piran, T.) 321-338 (North-Holland, Amsterdam, 1983).
10. Rosi, L. A. & Zimmerman, R. L. *Astrophys. Space Sci.* **45**, 447-466 (1976).
11. Carr, B. J. *Astr. Astrophys.* **89**, 6-21 (1980).
12. Taylor, J. H. & McCulloch, P. M. *Ann. N.Y. Acad. Sci.* **336**, 442-446 (1980).
13. Kahn, K. & Penrose, R. *Nature* **229**, 185-187 (1971).
14. de Lapparent, V., Geller, M. & Huchra, J. P. *Astrophys. J.* **302**, L1-L5 (1986).
15. McBreen, B. & Metcalfe, L. *Nature* **330**, 348-350 (1987).
16. Chang, K. & Refsdal, S. *Nature* **282**, 561-564 (1979).
17. Schneider, P. & Weiss, A. *Astr. Astrophys.* **171**, 49-65 (1987).
18. Kayser, R., Refsdal, S. & Stabell, R. *Astr. Astrophys.* **166**, 36-52 (1986).
19. Paczynski, B. *Astrophys. J.* **301**, 503-516 (1986).
20. Maccacaro, T., Gioia, I. M., Maccagni, D. & Stoke, J. T. *Astrophys. J.* **284**, L23-L27 (1984).
21. Ostriker, J. P. & Vietri, M. *Nature* **318**, 446-448 (1985).
22. Angel, J. R. & Stockman, H. S. A. *Rev. Astr. Astrophys.* **18**, 321-362 (1980).
23. Paczynski, B. *Astrophys. J.* **308**, L43-L46 (1986).
24. Paczynski, B. *Astrophys. J.* **317**, L51-L55 (1987).
25. Mazats, E. P. & Golenskii, S. V. *Astrophys. Space Sci. Rev.* **1**, 205-266 (1981).
26. Mazats, E. P. *et al. Astrophys. Space Sci.* **80**, 3-143 (1981).
27. Atteia, J. L. *et al. Astrophys. J. Suppl. Ser.* **64**, 305-382 (1987).
28. Laros, J. G. *et al. Astrophys. J.* **290**, 728-734 (1985).
29. Atteia, J. L. *et al. Astr. Astrophys.* **152**, 174-176 (1985).
30. Barat, C. *et al. Astrophys. J.* **280**, 150-153 (1984).
31. Liang, E. P. *Astrophys. J.* **308**, L17-L20 (1986).
32. Metcalfe, L. & McBreen, B. *Astr. Astrophys.* **166**, L5-L7 (1986).
33. Norris, J. P. *et al. Astrophys. J.* **301**, 213-219 (1986).
34. Fegan, D. J., McBreen, B. & O'Sullivan, C. *Phys. Rev. Lett.* **51**, 2341-2344 (1983).
35. Mazats, E. P. *et al. Nature* **290**, 378-381 (1981).
36. Lasota, J. P. & Belli, B. M. *Nature* **304**, 139-141 (1983).
37. Cunningham, C. T. *Astrophys. J.* **202**, 788-802 (1975).
38. Thorne, K. S. *Astrophys. J.* **191**, 507-519 (1974).
39. Lingenfelter, R. E. & Ramaty, R. *AIP Conf. Proc.* **83**, 148-159 (1982).
40. Grindlay, J. E., Wright, E. L. & McCrosky, R. E. *Astrophys. J.* **192**, L113-L114 (1974).
41. Laros, J. G. *et al. Astrophys. J.* **286**, 681-691 (1984).
42. Noonan, T. W. *Astrophys. J.* **264**, 627-629 (1983).
43. Jennings, M. C. *Astrophys. J.* **295**, 51-64 (1985).
44. Yamagami, T. & Nishimura, J. *Astrophys. Space Sci.* **121**, 241-253 (1986).
45. Turner, E. L., Ostriker, J. P. & Gott III, J. R. *Astrophys. J.* **284**, 1-22 (1984).
46. Schneider, P. & Wagoner, R. V. *Astrophys. J.* **314**, 154-175 (1987).

Transport and optical studies of single crystals of the 80-K Bi-Sr-Ca-Cu-O superconductor

H. Takagi*, H. Eisaki*, S. Uchida*, A. Maeda†, S. Tajima†, K. Uchinokura† & S. Tanaka†

*Engineering Research Institute and †Department of Applied Physics, Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

One of the most challenging problems in solid-state physics is to elucidate the mechanism of the high-transition-temperature (high- T_c) superconductivity in cuprates. Two classes of cuprates, $(\text{La, Sr})_2\text{CuO}_4$ (LSCO) and $\text{YBa}_2\text{Cu}_3\text{O}_{7-y}$ (YBCO)^{1,2}, are now well established as high- T_c superconductors. Several characteristic features common to LSCO and YBCO have provided the basis for theoretical models of high- T_c superconductivity³. Various properties, in particular the transport properties, exhibit large anisotropy owing to the layered structure. The dominant charge carriers are holes, which are supplied in the Cu-O plane by the substitution of La with Sr in La_2CuO_4 or by adding oxygen to $\text{YBa}_2\text{Cu}_3\text{O}_6$. Recently, Maeda *et al.*⁴ announced the discovery of new superconducting oxides in the system Bi-Sr-Ca-Cu-O, with T_c above 100 K. Their materials exhibit the superconducting transition in two steps, at ~ 80 K and 105 K, indicating the presence of two superconducting phases. Although the detailed crystal structure is not known yet, several groups have reported that at least the 80-K superconductor has the formula $\text{Bi}_2(\text{Sr, Ca})_3\text{Cu}_2\text{O}_y$ ^{5,6}. Transmission electron microscopy reveals that the material possesses a layered structure similar to those of LSCO and YBCO, with Cu-O layers isolated between Bi-O layers⁷. Here we report the synthesis of single crystals of Bi-Sr-Ca-Cu-O with $T_c = 80$ K; and measurements of their basic properties. The similarity of the measured properties to those of LSCO and YBCO indicates the importance in the new materials, as in their predecessors, of two-dimensional Cu-O layers.

Single crystals of Bi-Sr-Ca-Cu-O were grown in an Al_2O_3 crucible by slowly cooling a Bi_2O_3 - SrCO_3 - CaCO_3 - CuO solution from 1,200 to 800 °C at 2 °C hr⁻¹ in air. The starting composition

of the solution was $\frac{1}{2}\text{Bi}_2\text{O}_3$: SrCO_3 : CaCO_3 : $\text{CuO} = 1:1:1:2$. Many shiny black blocks were found inside the solidified solution; these crystal aggregates yielded rectangular platy crystals with typical dimensions $1.5 \times 0.7 \times 0.1$ mm³.

The X-ray analysis shown in Fig. 1 indicates that our crystals are orthorhombic, with lattice parameters close to those reported on powdered samples: $a = 5.453$ (3) Å, $b = 27.26$ (4) Å, $c = 31.04$ (3) Å, the c -axis being perpendicular to the surface of the thin plate. A fivefold superstructure is present along the b -axis. No trace of twins, however, was observed in the oscillating crystal photograph.

The chemical composition of the crystals was roughly estimated by an electron probe micro-analyser (EPMA). A preliminary analysis indicates that the Bi: Sr: Ca: Cu ratio is almost the same from crystal to crystal, approximately 1:0.7:0.5:1. The (Sr, Ca) content is less than expected from the stoichiometry proposed in refs 5 and 6, but as the result of the EPMA analysis is subject to some uncertainty, we will denote the chemical composition of our crystal as $\text{Bi}_2(\text{Sr, Ca})_3\text{Cu}_2\text{O}_y$. More precise analysis of the composition is now under way.

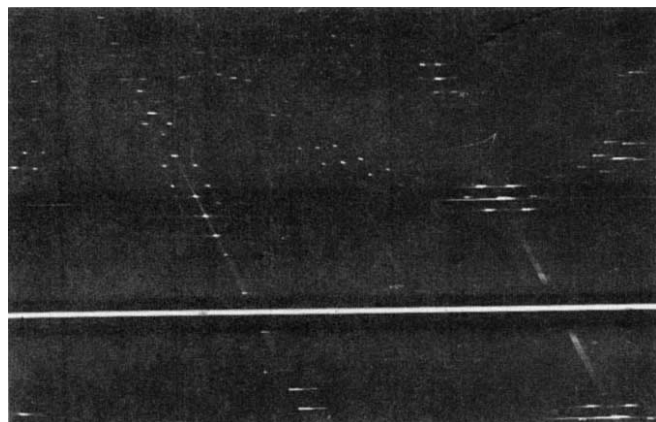


Fig. 1 Weissenberg X-ray diffraction pattern of the $(0, k, l)$ reflection of single-crystal $\text{Bi}_2(\text{Sr, Ca})_3\text{Cu}_2\text{O}_y$. The lattice constants in the text were determined by four-circle goniometer.

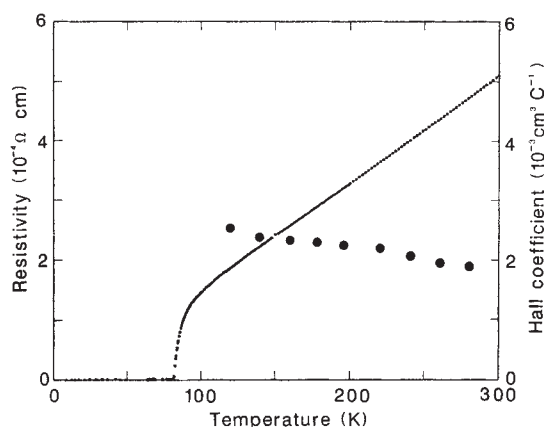


Fig. 2 Temperature dependence of the resistivity (curve) and Hall coefficient (points) for a $\text{Bi}_2(\text{Sr}, \text{Ca})_3\text{Cu}_2\text{O}_y$ single crystal. The current was applied parallel to the a -axis. A magnetic field of 18 kOe was applied parallel to the c -axis for the Hall coefficient measurement.

As-grown crystals were already 80-K superconductors and annealing in an oxygen atmosphere did not markedly improve the superconducting characteristics. There was often a decrease in resistivity at ~ 105 K, but the strength of the Meissner signal measured by a SQUID magnetometer indicated that there was $< 5\%$ of the 105-K phase. To avoid complications arising from the coexistence of two superconducting phases, crystals without any trace of 105-K superconductivity were selected for the present work.

Figure 2 shows the temperature dependence of the resistivity with the current j parallel to the a -axis. Electrical resistivity and Hall coefficient were measured using the conventional six-probe technique. The substantial magnitude of the Meissner signal measured by a d.c. SQUID magnetometer—30% of the ideal value—guarantees the bulk nature of the superconductivity with transition temperature $T_c \approx 80$ K. For all of the crystals investigated so far, the resistivity is $\sim 500 \mu\Omega \text{ cm}$ at 300 K, the same order of magnitude as in single-crystal LSCO and YBCO⁸, and shows almost linear dependence on temperature.

From the proposed layered structure, we can expect substantial anisotropy in the electronic states, as in LSCO and YBCO. This is confirmed by our measurements of anisotropy in the upper critical field (H_{c2}). Figure 3 shows the resistive superconducting transition in the presence of magnetic fields (H) of varying strength. When the magnetic field is applied, a significant tail appears in the curve, possibly due to variations of composition in the crystal. In spite of the broad superconducting transition in the applied fields, a large anisotropy is clearly seen (compare Fig. 3a, b with Fig. 3c). The transition temperature is lowered much more when the magnetic field is applied parallel to the c -axis, that is, perpendicular to the (Sr, Ca)-Cu-O layers sandwiched between Bi-O layers. Although the formation of a fivefold superstructure along the b -axis was observed, no significant difference is seen between the $H \parallel a$ and $H \parallel b$ cases. If we tentatively define the transition temperature as the midpoint of the transition, the initial slope of the upper critical field is estimated to be 4.5, 5.0 and 0.5 T K^{-1} for $H \parallel a$, b and c , respectively. The anisotropy in the initial slope (comparing the values for H parallel and perpendicular to the a - b plane) is ~ 10 —comparable to or larger than the values reported for LSCO and YBCO⁹. These results indicate that the conductivity in the plane is much larger, probably because the conduction mainly occurs through the Cu-O network, as in LSCO and YBCO.

Figure 2 shows the temperature dependence of the Hall coefficient, measured with $j \parallel a$ and $H \parallel c$. The Hall coefficient is positive in sign and increases slightly with decreasing temperature. Its magnitude at 300 K is $\sim 2 \times 10^{-3} \text{ cm}^3 \text{ C}^{-1}$. It

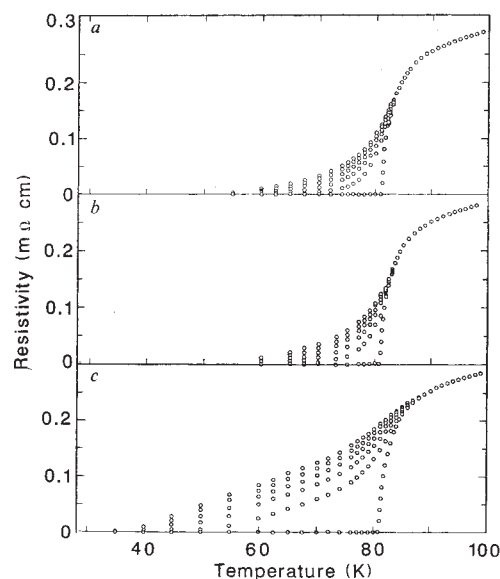


Fig. 3 Temperature dependence of the resistivity in an applied magnetic field for a $\text{Bi}_2(\text{Sr}, \text{Ca})_3\text{Cu}_2\text{O}_y$ single crystal. The magnetic field was applied parallel to the a , b and c axes, in parts a, b and c, respectively. The six curves in each figure correspond to $H = 5, 4, 3, 2, 1$ and 0 T, respectively, from top to bottom.

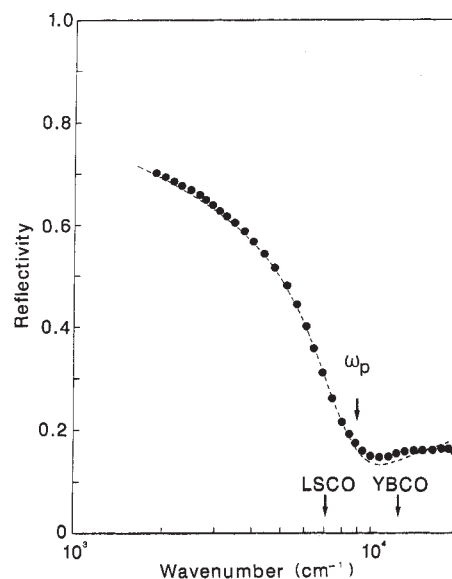


Fig. 4 Optical reflectivity spectrum of a $\text{Bi}_2(\text{Sr}, \text{Ca})_3\text{Cu}_2\text{O}_y$ single crystal with the electric field parallel to the a - b plane. The points are the experimental data; the dashed line shows a fit to the Drude law with plasma frequency $\omega_p = 8,600 \text{ cm}^{-1}$, damping $\gamma = 5,500 \text{ cm}^{-1}$ and optical dielectric constant $\epsilon_{op} = 7.5$. The plasma energies of LSCO and YBCO are shown by arrows for comparison.

corresponds to a hole concentration (n) of $3 \times 10^{21} \text{ cm}^{-3}$ (0.35 holes per Cu atom), if we simply apply the relationship $R_H = 1/ne$. The result indicates that the conduction is dominated by a relatively low density of holes, as in LSCO and YBCO¹⁰. In LSCO and YBCO, the mean valence of Cu is a little larger than $2+$, and it has been pointed out that there is a close relationship between the number of holes per Cu atom and the mean valence of Cu¹¹. If this were also true for the present material, the mean valence of Cu would be 2.35, which corresponds to the composition $\text{Bi}_2(\text{Sr}, \text{Ca})_3\text{Cu}_2\text{O}_{8.35}$.

Another consequence of the low carrier density can be seen in the optical reflectivity spectrum measured with the electric

field parallel to the plane (Fig. 4). A plasma reflectivity edge is seen in the infrared region, and the spectrum fits the Drude formula, as in the case of LSCO and YBCO. The relatively low plasma frequency as well as the large Hall coefficient indicate the low carrier concentration. For comparison, the plasma frequencies of LSCO and YBCO¹² are shown by the arrows in Fig. 4. Note that the plasma energy increases in the sequence LSCO, Bi₂(Sr, Ca)₃Cu₂O_y and YBCO, in order of increasing T_c . As the plasma frequency increases with increasing carrier concentration, the result is suggestive of the correlation between T_c and hole concentration.

The superconducting oxide Bi₂(Sr, Ca)₃Cu₂O_y with $T_c = 80$ K exhibits strikingly similar properties to the previous lanthanide-based cuprates LSCO and YBCO, namely the linear dependence of resistivity on temperature, the anisotropic upper critical field and the positive Hall coefficient. The proposed structure of this material⁵ contains neither lanthanide layers, nor Cu–O chains. We therefore conclude that the normal-state properties, and possibly the superconductivity of Bi₂(Sr, Ca)₃Cu₂O_y is primarily governed by holes in two-dimensional Cu–O layers, as in the case of the previous high- T_c oxides.

We thank T. Wada of Matsushita Co. Ltd for performing EPMA analysis and discussion, and T. Hasegawa, K. Kishio, K. Kitazawa, K. Fueki of Tokyo University and Y. Maeno of Hiroshima University for stimulating discussion.

Received 23 February; accepted 7 March 1988.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 193–198 (1986).
2. Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908–910 (1987).
3. Rice, T. M. *Z. Phys.* **B67**, 141–144 (1987).
4. Maeda, H., Tanaka, Y., Fukutomi, M. & Asano, T. *Jap. J. appl. Phys.* (in the press).
5. Tarascon, J. M. *et al. Phys. Rev. B* (submitted).
6. Takayama-Muromachi, E. *et al. Jap. J. appl. Phys.* (submitted).
7. Matsui, Y. *et al. Jap. J. appl. Phys.* (submitted).
8. Tozer, S. W., Kleinsner, A. W., Penny, T., Kaiser, D. & Holtzberg, F. *Phys. Rev. Lett.* **59**, 1768–1770 (1987).
9. Iye, Y., Tamegai, T., Takeya, H. & Takeji, H. *Jap. J. appl. Phys.* **26**, L1057–L1059 (1987).
10. Takagi, H., Uchida, S., Iwabuchi, H., Tajima, S. & Tanaka, S. *Jap. J. appl. Phys.* (in the press).
11. Ong, N. P. *et al. Phys. Rev. B* **35**, 8807 (1987).
12. Tajima, S. *et al. Mod. Phys. Lett.* (in the press).

Significance of plane versus chain sites in high-temperature oxide superconductors

Gang Xiao, M. Z. Cieplak, D. Musser, A. Gavrin, F. H. Streitz, C. L. Chien, J. J. Rhyne* & J. A. Gotaas*

Department of Physics and Astronomy,
The Johns Hopkins University, Baltimore, Maryland 21218, USA
* Institute for Materials Science and Engineering,
National Bureau of Standards, Gaithersburg, Maryland 20899, USA

One of the outstanding questions concerning the high-temperature superconductor YBa₂Cu₃O₇ is the relative importance of the CuO₂ planes (Cu(2) site) and the CuO chains (Cu(1) site). Many theories^{1,2} have stressed the importance of the 3d holes at the Cu sites, which provide an antiferromagnetic correlated background necessary for the coupling of the superconducting electrons. We have substituted Zn(3d¹⁰4s²) and Ga(3d¹⁰4s²4p¹) in the Cu(2) and Cu(1) sites, respectively, while maintaining the oxygen stoichiometry at 7. In the valence states of 2+ and 3+ respectively, the 3d bands of both elements are completely full, without the complication of a magnetic moment. Here we discuss the structural changes that accompany Zn and Ga substitution, as deduced from X-ray and neutron diffraction, and correlate these changes with the superconducting properties. We show that the integrity of the planes is much more important than that of the chains in sustaining high- T_c superconductivity.

Table 1 Structural parameters of YBa₂(Cu_{0.94}Zn_{0.06})₃O_{7-δ} determined by neutron diffraction at 295 K

	$a = 3.8316(2) \text{ \AA}$			$R_N = 5.58$	
	$b = 3.8899(2) \text{ \AA}$			$R_{wp} = 12.88$	
	$c = 11.6674(6) \text{ \AA}$			$R_E = 6.29$	
	$V = 173.90(2) \text{ \AA}^3$			$\chi = 2.05$	
	$(b-a)/(b+a) = 7.55(6) \times 10^{-3}$			$\delta = 0.20(6)$	
Site	x	y	z	B	N
Y	0.5	0.5	0.5	0.55 (9)	1.0 (f)
Ba	0.5	0.5	0.1836 (4)	1.2 (1)	2.0 (f)
Cu(1)	0.0	0.0	0.0	1.1 (1)	1.00 (5)
Zn(1)	0.0	0.0	0.0	1.1 (1)	0.00 (5)
Cu(2)	0.0	0.0	0.3547 (3)	0.47 (8)	1.80 (8)
Zn(2)	0.0	0.0	0.3547 (3)	0.47 (8)	0.20 (8)
O(1)	0.0	0.0	0.1590 (4)	0.9 (1)	2.0 (f)
O(2)	0.5	0.0	0.3771 (4)	0.9 (1)	2.0 (f)
O(3)	0.0	0.5	0.3767 (4)	0.3 (1)	1.90 (3)
O(4)	0.0	0.5	0.0	0.09 (1) [β_a]	0.84 (2)
	—	—	—	0.034 (8) [β_b]	
	—	—	—	0.004 (8) [β_c]	
O(5)	0.5	0.0	0.0	0.6 (f)	0.06 (1)
Comparison of alternate constrained-fit models					
Constrained occupancy				R_N	R_{wp}
Zn(1)=0.0, Zn(2)=0.18				5.58	12.88
Zn(1)=0.18, Zn(2)=0				5.62	12.94
Zn(1)=0.66, Zn(2)=0.12				5.58	12.89

Numbers in parentheses represent the statistical errors in the last decimal place, and 'f' indicates that the parameter was held constant in value. The site notation is that given in ref. 13. The total occupancy (Zn plus Cu) of the Cu(1) and Cu(2) sites was constrained to 1 and 2 atoms, respectively. The thermal parameter B for the orthorhombic O(4) (chain) site and for the O(1) site was refined with an anisotropic thermal ellipsoid of the form $\exp(-h^2\beta_a - k^2\beta_b - l^2\beta_c)$ and illustrates the significantly large mean square amplitude along the a -axis relative to the b and c directions. R_N , R_{wp} , R_E are the nuclear, weighted-profile and expected R -factor of fit respectively. N is the occupancy number.

Copper is a crucial element in the oxide superconductors. A number of groups have substituted the Cu sites by other 3d transition elements (from Ti to Zn)³⁻¹⁰ and metalloid elements (such as Al)^{6,11,12}, to elucidate the role of the Cu site. Although it has been shown that some of the dopants (Co, Ni and Fe) preferentially substitute either the Cu(1) or Cu(2) sites, the large magnetic moments of the dopants severely complicate the interpretation of the superconducting behaviour. Al was found by X-ray diffraction to substitute the Cu(1) site¹¹, but the T_c (superconducting transition temperature) data in the Al-doped samples remain controversial^{6,11,12}. The ideal dopants should be non-magnetic, with electronic configurations similar to that of Cu, and should preferentially substitute either the Cu(1) or Cu(2) sites. Zn and Ga, two elements next to Cu in the periodic table, are excellent candidates.

The samples were made by using a solid-state reaction method, as described in ref. 3. The structures were determined using an automated X-ray powder diffractometer. The diffraction peaks were fitted by modified gaussians, and the lattice parameters were determined by fitting the positions of at least 18 diffraction peaks with an accuracy better than 0.05%. Such procedures are necessary to observe the detailed structural changes. The determined lattice parameters are shown in Fig. 1. All of the YBa₂(Cu_{1-x}Zn_x)₃O₇ samples remain orthorhombic, with a large distortion $((b-a)/(b+a))$, but the structure of the YBa₂(Cu_{1-x}Ga_x)₃O₇ system is very sensitive to the doping and transforms rapidly to the tetragonal structure. Figure 2 shows the diffraction patterns of YBa₂(Cu_{1-x}Ga_x)₃O₇ in the range $46^\circ \leq 2\theta \leq 48^\circ$; the (200), (020) and (006) peaks, relating to the lattice parameters a , b and c respectively, are marked. The values of b and a rapidly approach each other on doping with Ga. At

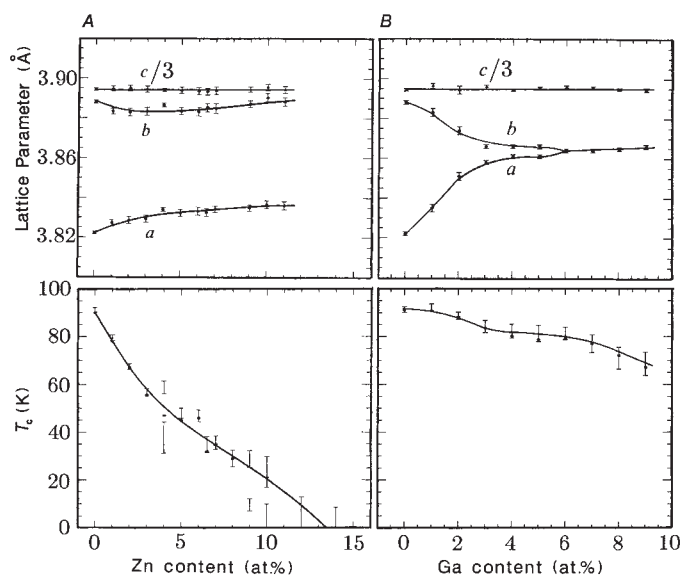


Fig. 1 Lattice parameters (a , b , $c/3$) and superconducting transition temperature (T_c) as functions of dopant content for YBa₂(Cu_{1-x}Zn_x)₃O₇ (A) and YBa₂(Cu_{1-x}Ga_x)₃O₇ (B). The filled dots are T_c s obtained from magnetization measurements at an applied field of 30 Oe, and the bars represent the resistive transition (90–10% of the resistivity drop at the onset).

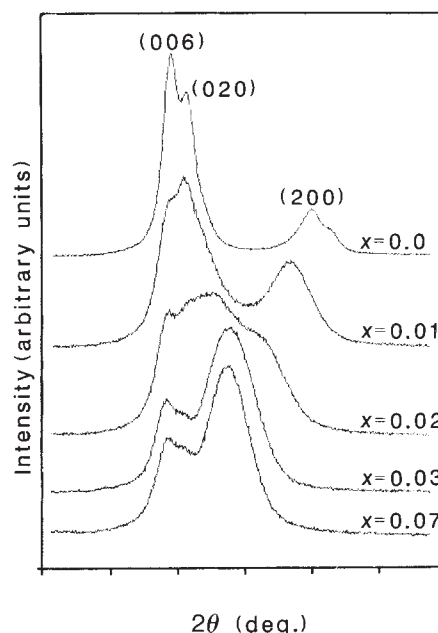


Fig. 2 X-ray diffraction patterns of YBa₂(Cu_{1-x}Ga_x)₃O₇ in the range $46^\circ \leq 2\theta \leq 48^\circ$, for five values of the gallium content, x . The (006), (020) and (200) peaks are related to the lattice parameters c , d and a , respectively.

$x = 0.03$, the (200) and (020) peaks nearly coincide, yielding a single but broadened peak. Careful fitting of at least 18 diffraction peaks indicates that the structure does not become fully tetragonal until $x = 0.06$.

The structures of two representative samples, YBa₂(Cu_{0.94}Zn_{0.06})₃O₇ and YBa₂(Cu_{0.92}Ga_{0.08})₃O₇, have also been measured by using neutron diffraction at the National Bureau of Standards reactor. The scattering intensities were refined using a modified Rietveld analysis procedure, and the essential results are summarized in Table 1 using the site notation of ref. 13. This technique is very sensitive to both oxygen site occupancies and small orthorhombic distortions, the latter sensitivity arising from the combination of high intrinsic spectrometer resolution and the well-defined line shape of the neutron peaks. Both samples exhibited total oxygen contents near seven: $O = 6.80(6)$ and $7.01(9)$ for the Zn and Ga samples, respectively. Thus, the sharply reduced orthorhombic distortion of the Ga-doped samples is not driven by oxygen deficiency.

From the neutron refinement, Zn was found to substitute into only the Cu(2) site, which does not lead to a disordering of the b -axis CuO chains. The structure remains orthorhombic, with essentially the same axis distortion as in the undoped YBa₂Cu₃O₇ with equivalent O(4) site occupancy. The refined occupancy for the four oxygen sites is given in Table 1. To test the validity of the conclusion that the Zn occupies only the Cu(2) site, three additional refinements were made with constrained Zn (and Cu) occupancies, as shown at the bottom of Table 1. The results were compared to the result of the 'free' refinement (Table 1, centre) using a statistical significance test (equivalent to the Hamilton test). The results show that the Cu(2)-only substitution model is superior to the Cu(1)-only substitution model at the 95% confidence level (slope of the regression line $= 0.857 \pm 0.703$). A $\frac{1}{2}:\frac{2}{3}$ stoichiometric distribution of Zn on Cu(1) and Cu(2) sites does not significantly degrade the R_{wp} factor, nor is it definitively preferred by the significance test. But this distribution is not a stable occupancy when the populations are allowed to refine independently, and thus does not represent the least-squares minimum for the structure function. The replacement of Cu by Zn on the Cu(2) site only is

therefore indicated as the correct configuration.

In the 8% Ga-doped sample, the orthorhombic distortion is quite small, but can be validly determined within the neutron instrument resolution. The ratio $(b-a)/(b+a)$ is $1.02(8) \times 10^{-3}$, which is consistent with the X-ray values of Fig. 1 of lower statistical precision. Using the $Pmmm$ orthorhombic refinement, the total oxygen content was $7.01(9)$ atoms and O(4) and O(5) site occupancies were found to be $0.66(5)$ and $0.35(4)$ atoms respectively, indicating that considerable disorder is induced on the chain-site atoms. In the limit of vanishing orthorhombic distortion, the structure is tetragonal (space group $P4/mmm$), the O(4) and O(5) sites (also O(2) and O(3)) become indistinguishable, and the chain structure is fully disordered. The scattering pattern of the Ga sample was also obtained at 10 K to ascertain if increased orthorhombic distortion was present below T_c . The 10-K value $(b-a)/(b+a) = 0.80(7) \times 10^{-3}$ differs little from that at room temperature. Unfortunately, the neutron scattering amplitudes of Ga ($b = 0.729$) and Cu ($b = 0.772$) are too similar to allow a definitive determination of the site substitution at the 8% Ga level. The remaining refinement parameters (z coordinates and so on) were not significantly different from those quoted in Table 1 for the Zn and are thus not listed. Note, however, that the appreciable O(4)–O(5) oxygen disorder and the very rapid transition towards a tetragonal structure on doping with Ga (see Fig. 1) strongly suggest that Ga substitutes preferentially on the Cu(1) sites, effectively destroying the chain ordering.

The superconducting transition temperatures were determined by both four-probe resistivity and SQUID magnetometry on precision-cut samples. The results are shown in Fig. 1, with good agreement between the two measurements. In the Zn samples, T_c reduces rapidly with Zn content, vanishing at $x_{Zn} \approx 0.13$. In the Ga samples, T_c is reduced only slightly by doping. At $x_{Ga} = 0.06$, where the structure becomes tetragonal, T_c remains as high as 81 K, and thus the high T_c observed is intrinsic to the tetragonal phase. Furthermore, the magnitude of the Meissner effect in the tetragonal phase remains almost the same as that on the orthorhombic YBa₂Cu₃O₇, indicating bulk superconductivity.

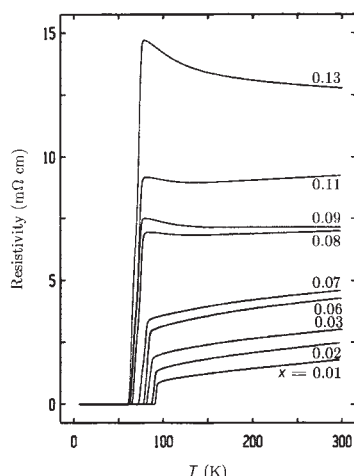


Fig. 3 Temperature dependence of resistivity for $\text{YBa}_2(\text{Cu}_{1-x}\text{Ga}_x)_3\text{O}_7$.

Our T_c data for Zn-doped samples are in agreement with results from other groups⁵⁻⁸. For example, Takayama-Muromachi *et al.*⁵ have shown that T_c drops to zero at ~13 atm% Zn ($x = 0.4$ in their notation), which is very close to what we have observed.

As Al is similar to Ga in charge and ionic radius, we would expect the Al-doped and Ga-doped systems to be similar. Both show an orthorhombic to tetragonal transition^{6,11}. Siegrist *et al.*¹¹ have shown that Al occupies the Cu(1) site, supporting our claim that Ga substitutes the chain site. The T_c s of the ceramic samples (at 1.7% and 3.3% Al) are similar to those of our Ga-doped samples, but there is a sharp suppression in T_c in the single-crystal samples above 3.3% doping. Such behaviour is not supported by the work of other groups^{6,12}, in which T_c is found to remain high (80 K) up to high (10%) Al doping levels, as for the Ga-doped samples studied here.

Figure 3 shows the temperature dependence of the resistivity of the Ga-doped samples. In the range $0 \leq x \leq 0.07$, the temperature dependence of resistivity in the normal state is very similar, being metallic in nature with a positive temperature coefficient of resistivity (TCR). The resistivity increases nearly in proportion to x . This behaviour is probably due to impurity scattering, which is generally not temperature dependent, but only increases the absolute resistivity. Above $x = 0.07$, the resistivity behaves very differently: in addition to increased normal-state resistivity, the samples no longer have a positive TCR, and there is always an upturn in resistivity before the superconducting transition. Note also the large increase of resistivity between $x = 0.07$ and $x = 0.08$. This change correlates closely with the loss of orthorhombic symmetry.

The charge state of Zn is 2+, the same as that of the Cu(2) site. According to a charge-balance argument the Cu(1) site should be in the 3+ state, which is also the charge state of Ga. The full oxygen stoichiometry is assured by the similarity between the dopant charge states and those of the Cu sites.

Zn and Ga ions both have a $3d^{10}$ configuration. On substituting a Cu site, the local $3d$ holes will be eliminated. As the oxygen contents of both systems remain at 7, the sharply different behaviours of the Zn- and Ga-doped samples must be due to their preferential substitution of the Cu(2) and Cu(1) sites, respectively. The results indicate that the $3d$ holes on the Cu(2) sites are important for high T_c ; the elimination of these holes by the introduction of Zn is deleterious to the superconductivity. On the other hand, the $3d$ holes on the Cu(1) sites play a much less important part; the filling of these holes by Ga ions affects the superconductivity only slightly. We therefore conclude that the CuO_2 planes play a much more decisive role than the CuO chains in sustaining the high- T_c superconductivity.

Received 24 December 1987; accepted 17 February 1988.

1. Anderson, P. W. *Science* **235**, 1196-1198 (1987).
2. Emery, V. J. *Phys. Rev. Lett.* **58**, 2794-2797 (1987).
3. Xiao, G., Streitz, F. H., Gavrin, A., Du, Y. W. & Chien, C. L. *Phys. Rev.* **B35**, 8728-8784 (1987).
4. Maeno, Y. *et al.* *Nature* **328**, 512-514 (1987).
5. Takayama-Muromachi, E., Uchida, Y. & Kato, K. *Jap. J. appl. Phys.* (in the press).
6. Tarascon, J. M. *et al.* *Phys. Rev. B* (submitted).
7. Jee, C. *et al.* *Proc. 1987 MRS Meet. High T_c Superconductivity* (in the press).
8. Oseroff, S. B. *et al.* *Solid St. Commun.* **64**, 241-245 (1987).
9. Kistenmacher, T. J. *et al.* *Phys. Rev. B* **36**, 8877-8880 (1987).
10. Miceli, P. F. *et al.* *Phys. Rev. B* (submitted).
11. Siegrist, T. *et al.* *Phys. Rev. B* **36**, 8365-8368 (1987).
12. Kirby, P. B. *et al.* *Phys. Rev. B* **36**, 8315-8319 (1987).
13. Beech, F. *et al.* *Phys. Rev. B* **35**, 8778-8781 (1987).

Evaluation of the Montsouris series of ozone measurements made in the nineteenth century

Andreas Volz & Dieter Kley

Institut für Chemie 2: Chemie der Belasteten Atmosphäre
Kernforschungsanlage Jülich, D-5170 Jülich, FRG

There is growing evidence that ozone levels in the lower troposphere over the continents of the Northern Hemisphere have been increasing during the past decades^{1,2}. Questions regarding pre-industrial or 'background' ozone concentrations have led to the search for data from the early days of ozone monitoring, during the second half of the last century. Unfortunately, most measurements were then made using Schönbein test paper, giving only semi-quantitative information due to poor standardization and the influence of humidity and wind speed on its sensitivity³⁻⁵. We have reinvestigated a set of ozone measurements gathered at the Observatoire de Montsouris, located on the outskirts of Paris, where a quantitative method was established in 1876⁶ and used continuously for 34 years. The evaluation of the technique, together with the analysis of nearly 3,000 of the original daily measurements that previously remained unnoticed in a statistical bulletin of the City of Paris⁷, provides conclusive evidence that ozone levels in central Europe 100 years ago averaged 10 p.p.b. and exhibited a seasonal variation, with a maximum during the spring months. Comparisons with modern data show that ozone levels in rural areas have more than doubled over the past century and that the tropospheric ozone budget is now strongly influenced by photochemical production due to increased levels of NO_x .

The method used at Montsouris was 'iodine catalysed oxidation of arsenite in neutral aqueous solution', see equation (1).

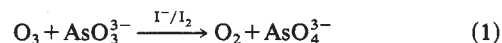
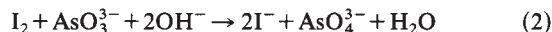


Figure 1 shows the experimental set up, consisting of a bubbler, a precisely calibrated gas meter and a water jet pump⁸. The interior part of the bubbler was made of platinum, with 20 upward facing holes at the bottom. Air entered the Pt tube directly, without the use of an inlet line. Every day the bubbler was prepared, mounted on a balcony of the observatory 5 m above the ground, and ambient air, at a flow rate of 1-2 l per min, was sampled for 24 h. After sampling, the remaining AsO_3^{3-} was titrated with a $5 \times 10^{-4} \text{ M}$ I_2 solution after the addition of $(\text{NH}_4)_2\text{CO}_3$ and starch (see equation (2)). An unexposed sample of the arsenite solution served as a reference.



Collection efficiency was shown to be 100% during the first year of operation through the use of two bubblers in series⁶. The AsO_3^{3-} in the second bubbler was never found to be depleted, a result which also demonstrated that the Pt-catalysed oxidation

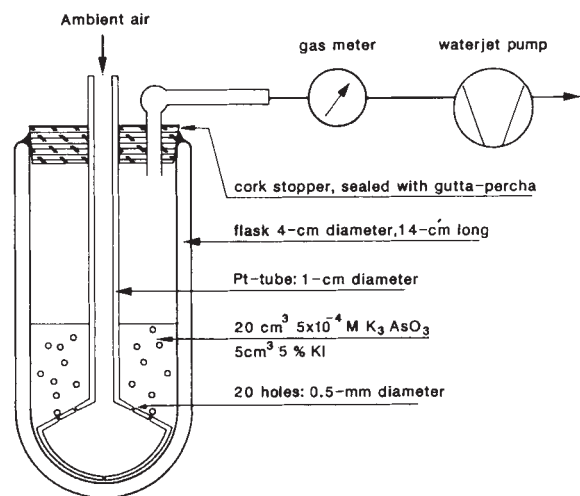


Fig. 1 Arrangement used at Montsouris for ozone measurements. The bubbler was mounted on a balcony ~5 m above the ground.

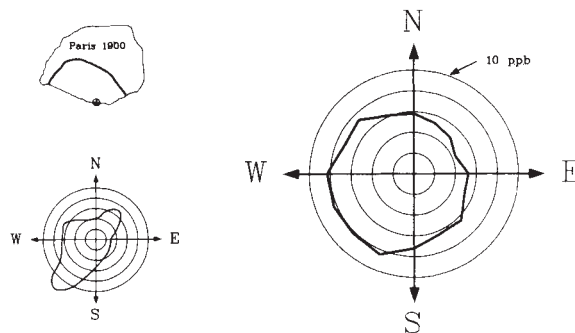


Fig. 2 Left panel: location of Montsouris (⊕) and average wind rose. Right panel: ozone rose of Montsouris derived from 3,000 daily measurements taken during 1876–86. Days with wind speeds of $<1 \text{ km h}^{-1}$ were discarded.

of AsO_3^{3-} by O_2 was unimportant during the 24-h sampling period. The method was known to be sensitive to other strong oxidants⁹ like H_2O_2 , and subject to negative interferences by reducing gases¹⁰. Note that, after the discovery of HCHO and SO_2 in the polluted air of Paris, only HCHO—but not SO_2 —was discussed in this context^{11,12}. Beginning in 1905, parallel experiments were conducted for the quantification of the reducing gases. For this purpose, an identical bubbler, equipped at the inlet with a 4-m natural rubber hose, was used. The latter was shown to destroy ozone completely¹⁰. According to these measurements, the interferences averaged 2.9 p.p.b. during the period 1905–1910 (details to be published).

We reconstructed the type of bubbler used at Montsouris and compared the method with the modern standard technique, namely ultraviolet absorption (details to be published). A constant amount of O_3 was generated by photolysis of O_2 in a flow of purified synthetic air. This flow was diverted for O_3 measurement both by ultraviolet absorption and by the Montsouris method. The flow rate through the bubbler was continuously monitored using a thermal mass flowmeter. Preparation of the bubbler and the titrations were performed as outlined by Albert-Lévy⁶.

Experiments were performed on O_3 concentrations ranging from 60 to 110 p.p.b. and at flow rates of 1–2.5 l per min. Due to the imprecision of the ultraviolet measurements (± 2 p.p.b.), higher concentrations than those found at Montsouris had to be used. The total amount of O_3 in each experiment, however, was kept similar to that at Montsouris ($\sim 1 \mu\text{M}$). The overall uncertainty, as calculated from the individual contributions of the O_3 measurement using ultraviolet absorption (± 2 p.p.b.), the flow rate ($\pm 1\%$) and the titration ($\pm 0.03 \mu\text{M}$), was 3–5%. The stoichiometry of the Montsouris method as determined from six individual experiments was found to average 1 ± 0.02 (total range: 0.96–1.04).

Additional tests were performed to determine the potential interferences. For this purpose, SO_2 and NO_2 were added to the gas flow from calibrated permeation tubes. H_2O_2 and HCHO were added to the bubbler in the form of aqueous solutions before, during and after the experiments. The results were as follows: H_2O_2 and NO_2 cause a positive interference, while SO_2 causes a negative interference with a stoichiometry of 1. HCHO, on the other hand, gives no significant response, even in amounts as large as $10 \mu\text{M}$. Therefore, SO_2 (and not HCHO) was the likely reducing species responsible for the interferences encountered at Montsouris. We also found (details to be published) that the method used at Montsouris between 1905 and 1910 may

have underestimated the interferences in the case of SO_2 by up to a factor of 2.

Therefore, we estimated the magnitude of the interferences from an analysis of the observed O_3 concentrations with respect to wind direction under the assumption that most of the SO_2 came from the nearby city of Paris which, at the turn of the century, had ~ 2.5 million inhabitants. For this purpose, the original daily O_3 observations, together with the required information on wind speed and direction, were retrieved for the years 1876–86 (except for 1881)⁷ and for the year 1905¹⁰. Altogether, about 3,000 individual daily ozone observations (except for 300 days with variable wind direction or wind speeds of $<1 \text{ km h}^{-1}$) were grouped into 16 wind sectors and averaged. The resulting ozone rose is shown in Fig. 2, along with the corresponding wind rose and a map of the location of Montsouris at the southern edge of Paris.

As Fig. 2 illustrates, uniform ozone concentrations of ~ 8 p.p.b. prevailed in the south-west sector, while concentrations were much lower in the presence of northeasterly winds. As there were no significant sources of pollution to the south-west of Montsouris, that sector was virtually free of interferences. On the other hand, the lower values in the north-east would reflect the interference of SO_2 acting as 'negative' ozone. By comparing the mean O_3 concentration from all wind directions to that of the south-west sector, an average interference of ~ 2 p.p.b. is obtained for the period 1876–86. An identical analysis for the year 1905 yields an O_3 concentration of 10 p.p.b. in the south-west sector and an average interference of 5 p.p.b.

Regarding interferences by oxidizing species, NO_2 concentrations in the south-west sector were certainly $\ll 1$ p.p.b., given the absence of motor traffic and power plants. More difficult to estimate are historic H_2O_2 concentrations, because this species, like O_3 , is not emitted directly into the atmosphere, but produced *in situ* by the recombination of perhydroxyl radicals (HO_2). According to reactions (3) to (8), HO_2 (and thus H_2O_2) is controlled by O_3 and NO_x . The lower O_3 concentrations at Montsouris would suggest lower H_2O_2 concentrations, as well. However, this effect was at least partially compensated for by the then lower NO_x concentrations. Consequently, H_2O_2 concentrations at Montsouris could have been similar to those found today, that is, between 0 and 2 p.p.b.¹³

As a result of our analysis, we applied the following correction to the observed O_3 concentrations^{10,14–19}: for interferences by reducing gases, namely SO_2 , a correction was made that increased in a linear fashion from 2 p.p.b. in 1881 to 5 p.p.b. in 1905. No correction was made for positive interferences, for example, H_2O_2 . The uncertainty of the corrected data is estimated to be ± 2 p.p.b. The annual mean O_3 concentrations of the Montsouris series are shown in Fig. 3. They range from 5

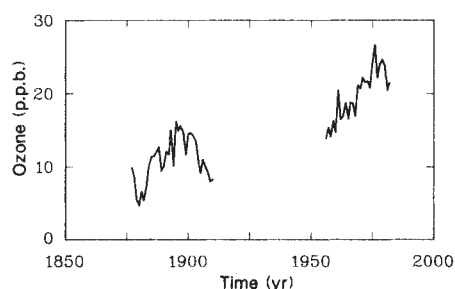


Fig. 3 Annual mean ozone concentrations at Montsouris (1876-1910) and at Arkona (1956-83). The Montsouris data are corrected for SO_2 interferences.

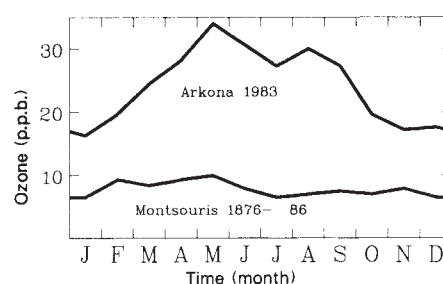


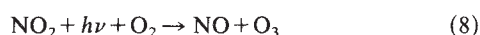
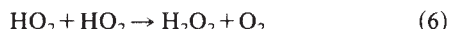
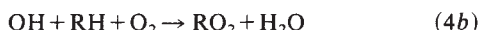
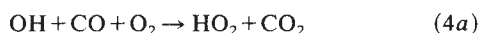
Fig. 4 Average seasonal variation of ozone at Montsouris (1876-86, south-west sector only) and at Arkona.

to 16 p.p.b., with the average for the entire 1876-1910 period being 11 p.p.b.

For comparison, O_3 measurements made since 1956 at Arkona²⁰ are also shown in Fig. 3. We selected this data set because Arkona is, due to its remote location on the island of Rügen in the Baltic Sea, almost unaffected by pollution, and because it has the longest record of modern O_3 measurements. The annual mean O_3 concentrations at Arkona increased from 15 p.p.b. in 1956 to 24 p.p.b. in 1983 at an average rate of 0.35 p.p.b. per year. At other rural monitoring stations in Europe, similar or even higher O_3 concentrations are observed today^{2,21-23}. Evidently, background O_3 concentrations have more than doubled since the beginning of this century.

The uncorrected Montsouris data show maximum O_3 concentrations in summer^{10,24}. We investigated the influence of SO_2 interferences on the seasonal variation by analysing the daily data from the period 1876-86 separately for the north-east and the south-west sectors and found that the average seasonal variation was dominated by the north-east sector, where extremely low values prevailed during winter. As this sector was strongly influenced by SO_2 interferences, the apparent decrease of O_3 was most likely a result of enhanced SO_2 concentrations during winter. For the reasons given above, the data most representative of the seasonal variation of O_3 at Montsouris are those from the south-west sector. These are displayed in Fig. 4 and show fairly constant values throughout the year, with a slight maximum in spring. For purposes of comparison, the seasonal variation found at Arkona²⁰ in 1983 is also shown in Fig. 4. At Arkona, O_3 exhibits a broad summer maximum. A similar seasonal pattern is found in most modern measurements made at northern mid-latitudes^{2,25-27}.

The extremely low absolute concentrations, as well as the seasonal variation of O_3 at Montsouris, point to much lower tropospheric NO_x concentrations than are found today; in the presence of NO_x , enhanced radiation during summer leads to an increase in O_3 through reactions (7) and (8).



Conversely, in the absence of NO_x , increased solar radiation leads to a net destruction of O_3 (reactions 3 to 6) in summer. In addition, transport of O_3 from the stratosphere is most efficient in spring²⁸. As a result, an O_3 maximum in early spring, as observed in the Montsouris data, is expected in the presence

of low NO_x concentrations. We may thus conclude that, contrary to the situation 100 years ago, ozone in the troposphere is today strongly influenced by photochemical *in situ* production due to increased levels of NO_x .

Received 5 October 1987; accepted 25 January 1988.

1. Warmbt, W. *Z. Met.* **29**, 24-31 (1979).
2. Atmannspacher, W., Hartmannsgruber, R. & Lang, P. *Met. Rdsch.* **37**, 193-199 (1984).
3. Fox, C. B. *Ozone and Antozone* (Churchill Publ., London, 1873).
4. Linvill, D. E., Hooker, W. J. & Olson, B. *Mon. Wea. Rev.* **108**, 1883-1891 (1980).
5. Kley, D., Volz, A. & Mülheims, F. *Proc. Adv. Res. Workshop Tropospheric Ozone Lillehammer, Norway, June 1-5, 1987* (ed. Isaksen, I. S. A.) 67-72 (Reidel, Dordrecht, 1987).
6. Albert-Lévy, Annuaire de l'Observ. de Montsouris 495-505 (Gauthier-Villars, Paris, 1878).
7. *Bull. hebdomadaire Statistique municipale Ville de Paris, République Française* (1876-1886).
8. Marié-Davy, H. *Bull. mens. Obs. Montsouris* **5**, 57-61 (1876).
9. Albert-Lévy, *Annls Obs. mun. Montsouris* **5**, 22 (1904).
10. Albert-Lévy, Henriot, H. & Bouyssy, *Annls Obs. mun. Montsouris* **6**, 18-22, 120-125, 212-216 and 315-321 (1905).
11. Henriot, H. *Annls Obs. mun. Montsouris* **4**, 63-65 and 295-298 (1908).
12. Henriot, H. *Annls Obs. mun. Montsouris* **7**, 424-426 (1906).
13. Heikes, B. G., Kok, G. L., Walega, J. G. & Lazrus, A. L. *J. geophys. Res.* **92**, 915-932 (1987).
14. Albert-Lévy *Annls Obs. mun. Montsouris* **1**, 97 (1900).
15. Albert-Lévy *Annls Obs. mun. Montsouris* **7**, 368 (1906).
16. Albert-Lévy *Annls Obs. mun. Montsouris* **8**, 289-291 (1907).
17. Miguel, P. *Annls Obs. mun.* **9**, 8, 242 (1908).
18. Miguel, P. *Annls Obs. mun.* **10**, 6, 228 (1909).
19. Miguel, P. *Annls Obs. mun.* **11**, 6-7 (1910).
20. Feister, U. & Warmbt, W. *J. atmos. Chem.* **5**, 1-21 (1987).
21. *Monat. Berichte Messnetz Umweltbundesamt Berlin FRG* (1983-1987).
22. Derwent, R. G. (ed.) *Ozone in the United Kingdom* 29-39 (Dept. Environment, London, 1987).
23. Guderian, R. (ed.) *Air Pollution by Photochemical Oxidants* (Springer, Berlin, 1985).
24. Bojkov, R. D. *J. Clim. appl. Met.* **25**, 343-352 (1986).
25. Logan, J. A. *J. geophys. Res.* **90**, 10463-10482 (1985).
26. Oltmans, S. J. & Komhyr, W. D. *J. geophys. Res.* **81**, 5359-5364 (1976).
27. Parrish, D. D. *et al. J. atmos. Chem.* **4**, 63-80 (1986).
28. Levy, H., II, Mahlman, J. D., Moxim, W. J. & Liu, S. C. *J. geophys. Res.* **90**, 3753-3772 (1985).

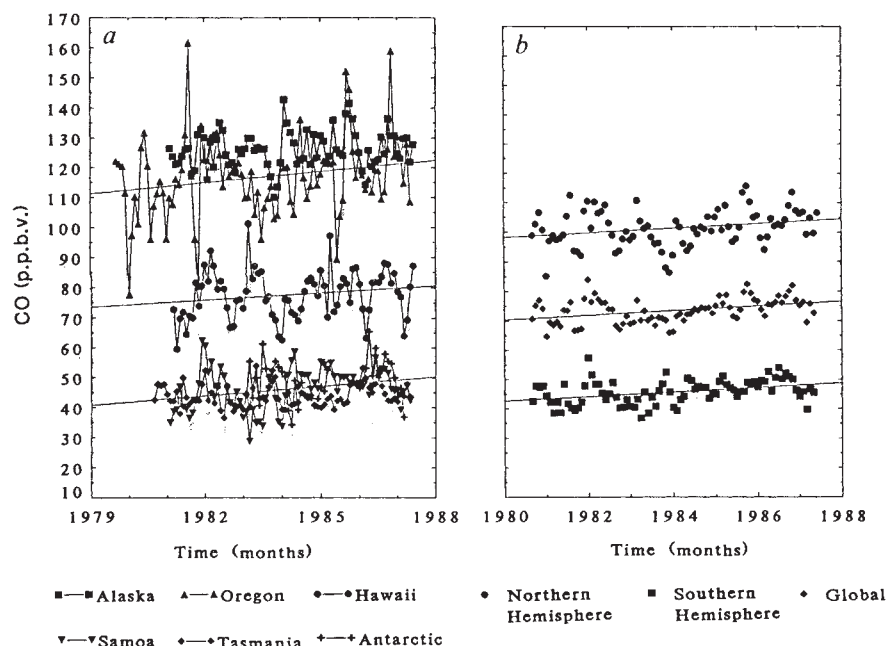
Carbon monoxide in the Earth's atmosphere: indications of a global increase

M. A. K. Khalil & R. A. Rasmussen

Institute of Atmospheric Sciences, Oregon Graduate Center, Beaverton, Oregon 97006, USA

Over half of the carbon monoxide in the atmosphere comes from human activities including motor traffic, other combustion of fossil fuels, and slash and burn agriculture¹⁻⁴. Additional anthropogenic sources include the burning of wood, savannah lands, and the oxidation of hydrocarbons including methane. Over the years these sources have increased gradually and may have already caused the concentrations of CO to double since pre-industrial times when human activities did not significantly affect the global cycles of CO and other trace gases. Increasing levels of CO can lead to an increase of tropospheric O_3 (refs 5,6) and a build-up of many other trace gases in the Earth's atmosphere, which may in turn cause

Fig. 1 *a*, The concentrations of carbon monoxide at six sites ranging in latitude from within the Arctic Circle to the South Pole. The average seasonal cycles have been subtracted from the data. The latitudinal variation of CO is apparent. *b*, Hemispheric and global averages of de-seasonalized CO concentrations. Concentrations are about 2.3 times higher in the Northern Hemisphere than in the Southern Hemisphere. The global average is shown in the middle.



widespread perturbations of tropospheric chemistry, global warming, and other climatic changes⁷. In a recent report⁸ to senior US Government officials the National Academies stated the urgent need to know the global distribution and trends of CO. During the past 6–8 years we have taken systematic measurements of CO at sites ranging from within the Arctic Circle to the South Pole. The rates of increase of the globally averaged concentration are between 0.8% and 1.4% per year depending on the statistical method used for estimating the trends. These increases may have gone on for much longer because more than half of the atmospheric CO now comes from anthropogenic sources. We find that the rates of increase are largest at mid-northern and tropical latitudes, where most of the sources are located.

Reaction with carbon monoxide is the principal mechanism for removing hydroxyl (OH) radicals from the global atmosphere. These radicals remove many manmade and natural trace gases from the troposphere, including almost all the hydrocarbons and many halocarbons. If the concentration of CO increases, tropospheric concentrations of OH may be depleted, causing even naturally occurring trace gases to build up in the atmosphere, and it may accelerate the increases of anthropogenic trace gases. For instance, the increase in methane concentration, which started more than a century ago, may be accelerated by decreasing levels of OH, and ultimately, more methane is likely to accumulate in the atmosphere if OH levels continue to decline. Calculations suggest that the concentrations of OH could be 20% lower now than a century ago, probably caused by the increase of CO (and CH₄)^{9–11}.

Because so much of the CO comes from human activities^{1–4} and the anthropogenic sources have been increasing, it has been suspected that the atmospheric concentration of CO may be increasing. There has been, however, little evidence for any such long-term global increase. Based on an analysis of spectroscopic plates from Jungfraujoch observatory in Switzerland, Rinsland and Levine¹² found that CO may have increased at ~2% per year between 1951 and 1985. Data from Zvenigorod, ~50 km from Moscow, although extremely variable, are consistent with an increase of ~2% per year between 1974 and 1982 (refs 13–15). In these studies there were data only at the beginning and ends of the periods studied. We reported an analysis of an intensive but short time series of CO measurements taken at Cape Meares, which showed a rapid increase of ~5% per year between 1979 and early 1983 (ref. 16). This fast increase was not representative of a long-term rate as we show here, but demonstrated that CO

was increasing in the non-urban troposphere. Here we report the trends of CO from systematic measurements taken during the past 6–8 years at seven sites ranging from the Arctic to the South Pole. The results demonstrate that the average tropospheric concentration of CO is increasing at between 0.8% and 1.4% per year, depending on the method used to estimate the trend, and the 90% confidence limits of the various estimates range between 0.5% and 2% per year. Moreover, the rates vary both spatially and temporally. It is not known whether these rates have persisted for a long time or are likely to be sustained. However, CO has probably been increasing over a long period because the sources have increased. We are reporting a continuing change in atmospheric composition although we can document and quantify the trends only over the period spanned by our measurements.

A major obstacle to estimating or even detecting trends in CO is the large spatial and temporal variability of its concentration. Much of the temporal variability is in the form of seasonal variations that repeat every year or two and the spatial variability is with latitude. Highest concentrations are observed in middle and high northern latitudes where most of the sources are located, and highest concentrations during the year are observed in winter, when the natural levels of OH, which remove CO from the troposphere, are at their minimum. When these systematic variations are accounted for, the remaining random variability is small enough for trends to be detected.

Our database consists of flask samples taken once a week at Barrow, Arkansas (72° N), Cape Meares, Oregon (45° N), Hawaii (Cape Kumukahi and Mauna Loa, 20° N), Samoa (12° S), Tasmania (42° S), Mawson¹⁷, Antarctica (72° S) and the South Pole, and about 50 automated measurements per day taken at Cape Meares.

To estimate the increasing trends at each site we first subtracted the average seasonal cycles at each site to obtain 'de-seasonalized' concentrations¹⁸. The average seasonal cycle is estimated as $\bar{\Delta}_{ij} = (1/N) \sum_{i=1}^N (C_{m_{ij}} - \bar{C}_{r_{ij}})$, where $C_{m_{ij}}$ is the measured average concentration of CO at a given site during month j ($j = 1, \dots, 12$) and year i ($i = 1, \dots, N$) when there are N years of data. $\bar{C}_{r_{ij}}$ is the corresponding concentration of CO at month i and year j obtained by a 12-month moving average. For the Barrow data a 24-month cycle was subtracted.

In addition to the site-by-site analysis of the data, we used the de-seasonalized data to form average monthly concentrations representative of each hemisphere and of the entire atmos-

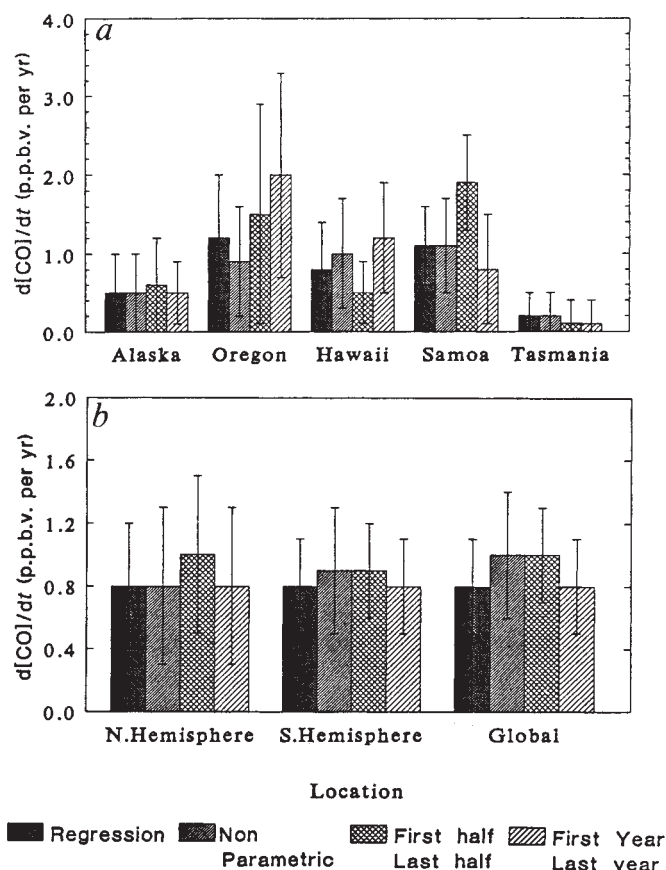


Fig. 2 *a*, The estimated rates of increase of CO at each site based on four methods of analysis discussed in the text. At all sites increasing trends are apparent ($\alpha = 0.1$ level). The largest increases appear to be at middle northern and tropical latitudes, where most of the sources are located. Lower trends are observed at high latitudes and middle southern latitudes. The first bar on the left for each site is from regression analysis, the second from non-parametric analysis of the slope, the third is from the comparison of the first half of the time series with the last half, and the last bar in each group is the increase estimated by comparing the average concentration during the first and last years of the time series. The error bars are 80% confidence limits. The lower limits are the lower estimates of the trends at the 10% level. *b*, Estimated trends based on the hemispherical and global average time series of Fig. 1*b*. The error bars are the 90% confidence limits of average CO concentrations. The rate of increase in per cent per year $= 100b/a$, where b is the rate in p.p.b.v. yr⁻¹ and a is the estimated concentration at the beginning of the time series. a (p.p.b.v.) = 125 (Barrow), 111 (C. Meares), 75 (Mauna Loa), 41 (Samoa), 43 (Tasmania), 97 (Northern Hemisphere), 42 (Southern Hemisphere), 70 (global average). The scale of concentration is based on our standards and may not be exactly the same as the scale used by others.

phere. We divided the atmosphere into four regions of equal volume and mass: 90–30° and 30–0° in each hemisphere. The average of the concentrations at Cape Meares and Barrow were taken to represent monthly average levels of CO at northern middle and high latitudes (90–30° N), average concentrations at Cape Kumukahi and Mauna Loa were taken to be representative of 30–0° N. Concentrations at Samoa represented 0–30° S, and the average of the concentrations at Tasmania, Mawson and the South Pole were taken to represent the region from 30–90° S. Hemispheric averages consisted of average monthly concentrations from the regions between 0–30° and 30–90°. The global average was calculated as the average of the concentrations in the two hemispheres. Although we have only a few sites in each hemisphere and we are not sure whether the averages are representative of the amount of CO in each hemisphere, the

averaged concentrations have much less variability and the increasing trends are more apparent. Figure 1*a* shows the de-seasonalized concentrations at each site, and Fig. 1*b* shows the hemispheric and global average concentrations.

We chose four methods to estimate the trends of CO in the de-seasonalized time-series data. The data were analysed by linear least squares and by non-parametric statistical methods^{19,20}. The third method consisted of dividing the data into two periods: the first and second halves. The average concentrations in the first (C_i) and second (C_f) halves were compared, the rate of increase was $b = 2(C_f - C_i)/T$, where T is the length of the time series (in years). The fourth method compared the average de-seasonalized concentrations during the first (C_i) and last (C_f) years when samples were collected. The rate of increase was $b = (C_f - C_i)/(T - 1)$. The last two methods are not usual for trend analysis, but they provide both estimates of the trend and critical tests of whether the trends are sustained over the period of observations.

The trend analyses were conducted in two groups: first for each site and then for the spatially averaged concentrations. Data from Cape Kumukahi, Mawson and the South Pole were insufficient for site-by-site trend analyses by our methods but were included in the hemispheric and global averages. The results are plotted in Fig. 2*a* and *b* corresponding to the data shown in Fig. 1*a* and *b*. The rates of increase are given in parts per 10⁹ (p.p.b.v.) per year and may be converted to percentages per year by using the estimated concentrations at the beginning of the time series.

At all sites where we had sufficient data for analysis all four methods show that the concentrations are increasing, but the rates are marginally significant ($\alpha = 0.1$) except at Tasmania, where the increase was found not to be statistically significant. In all calculations the four methods of analysis agreed on both the estimated trends and whether they are significant. At Cape Meares, where we have the most data, the trends were larger during the first part of the time series as we reported earlier¹⁶; in subsequent years, however, the concentrations declined and rose again, making the overall trend during the entire 8 years much smaller than observed during the first 3½ years. The trends seem to be largest at mid-northern and tropical latitudes, where most of the sources are located. At high northern latitudes and at the middle and high southern latitudes the trends become smaller and are as yet marginally significant or insignificant.

The hemispheric and global average concentrations more clearly show increasing trends which are significantly greater than zero using all four methods ($\alpha = 0.05$). However, it is apparent from our analysis that CO is probably not increasing at the same rate everywhere. For example, the increase of the average Southern Hemisphere concentration is mostly due to the strong trend at Samoa and moderated by the lack of similar trends at higher southern latitudes farther away from sources. We believe that we have barely detected an increasing global trend of atmospheric CO; accurate estimates of the long-term rates of increase will require a much longer series of systematic measurements.

We thank P. Turner for data management and analysis; D. Stearns and A. J. Crawford for maintaining sites and laboratory analyses; and P. J. Fraser of CSIRO (Australia) for sampling at Tasmania and providing data from Mawson. Funding for this work was provided in part by grants from the NSF (ATM 8414020), NASA (NAGW-280) and the Department of Energy (DE-FG06-85ER6031); additional support was provided by the Biospherics Research Corporation and the Andarz Co.

Received 30 September; accepted 31 December 1987.

- Seiler, W. *Tellus* **26**, 116–135 (1974).
- Logan, J. A., Prather, M. J., Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **86**, 7210–7254 (1981).
- National Academy of Sciences *Carbon Monoxide* (National Research Council, Washington, DC, 1977).

4. Freyer, H.-D. in *The Global Carbon Cycle* (eds Bolin, B. et al.) Ch. 4 (Wiley, New York, 1979).
5. Fishman, J. & Crutzen, P. J. *Nature* **274**, 855-858 (1978).
6. Fishman, J., Ramanathan, V., Crutzen, P. J. & Liu, S. C. *Nature* **282**, 818-820 (1979).
7. National Academy of Sciences *Changing Climate* (National Research Council, Washington, DC, 1983).
8. National Academy of Sciences *The National Academies Report to Senior Government Officials* 10-12 (National Research Council, Washington, DC, 1985).
9. Khalil, M. A. K. & Rasmussen, R. A. *Atmos. Envir.* **19**, 397-407 (1985).
10. Levine, J. S., Rinsland, C. P. & Tennille, G. M. *Nature* **318**, 254-257 (1985).
11. Thompson, A. M. & Cicerone, R. J. *Nature* **321**, 148-150 (1986).
12. Rinsland, C. P. & Levine, J. S. *Nature* **318**, 250-254 (1985).
13. Dvoryashina, E. V., Dianov-Klovov, V. I. & Yurganov, L. N. *Izvestia, USSR, Atmospheric and Oceanic Physics* **20**, 40-47 (1984).
14. Dvoryashina, E. V., Dianov-Klovov, V. I. & Yurganov, L. N. *Results of Carbon Monoxide Abundance Measurements, Zvenigorod, 1970-1982* preprint of the Academy of Sciences of the USSR, Moscow (1982).
15. Khalil, M. A. K. & Rasmussen, R. A. in *Environmental Impact of Natural Emissions* (ed. Aneja, V. P., 403-414 (APCA, Pittsburgh, 1984).
16. Khalil, M. A. K. & Rasmussen, R. A. *Science* **224**, 54-56 (1984).
17. Fraser, P. J., Hyson, H. P., Rasmussen, R. A., Crawford, A. J. & Khalil, M. A. K. *J. atmos. Chem.* **4**, 3-42 (1986).
18. Khalil, M. A. K. & Rasmussen, R. A. *Science* **232**, 56-58 (1986).
19. Hollander, M. & Wolfe, D. A. *Nonparametric Statistical Methods* (Wiley, New York, 1973).
20. Snedecor, G. W. & Cochran, W. G. *Statistical Methods* (Iowa State University Press, Ames, 1980).

Wet and dry deposition of Chernobyl releases

M. J. Clark* & F. B. Smith†

* National Radiological Protection Board, Chilton, Didcot, Oxon OX11 0RQ, UK

† Meteorological Office, Bracknell, Berks RG12 2SZ, UK

The passage of the Chernobyl plume over the United Kingdom in May 1986 led to the deposition of radionuclides on the ground by wet and dry deposition processes. Here we analyse rainfall during the passage of the plume and the published monitoring data obtained afterwards, and show that levels of deposited ^{137}Cs can be closely related to rainfall intercepting the plume. ^{137}Cs was present in the atmosphere mostly as particulate species with wet deposition mechanisms dominating. In contrast, ^{131}I was present as particulate and vapour phase material, and reported levels on grass and in cow's milk show that both wet and dry deposition mechanisms were important. ^{131}I on grass and in cow's milk therefore shows a different geographic pattern to ^{137}Cs , and is not so closely related to rainfall. We have calculated washout factors for locations where there are data on deposition, rainfall and air concentrations during the passage of the Chernobyl plume. From these factors and interpolated concentrations in the atmosphere, the total deposition of ^{137}Cs has been estimated at each of 4,000 rain gauge stations in the United Kingdom. The results are presented as deposition contours and have been compared with measurements in parts of the country. Estimates of the total deposition of ^{131}I and ^{137}Cs show that $\leq 1\%$ of the estimated total releases from Chernobyl were deposited on the United Kingdom.

The major releases of radioactive materials from the accident at the Chernobyl power plant lasted nearly ten days¹, starting at 01:23 local time on 26 April 1986 and virtually ceasing on 6 May. It is estimated that $\sim 2 \times 10^{18}$ Bq of fission and activation products were released over this period, and a significant proportion of the volatile species were transported over a wide area of Europe. Trajectories of the radioactive plume have been evaluated at many centres and are described elsewhere, for example in refs 2, 3 and 4. There was no widespread rainfall in the Ukraine and neighbouring areas of the Soviet Union which intercepted the radioactive plume from Chernobyl¹. This lack of rainfall meant that a significant fraction of the volatile fission products, such as ^{131}I , ^{132}Te and ^{137}Cs , remained in the atmosphere until interception by rainfall in other parts of Europe led to appreciable deposition on the ground⁵. Rainfall in the United Kingdom occurring at the time of passage of the Chernobyl plume contained elevated levels of radionuclides, as shown by measurements of rainwater samples and of outdoor gamma

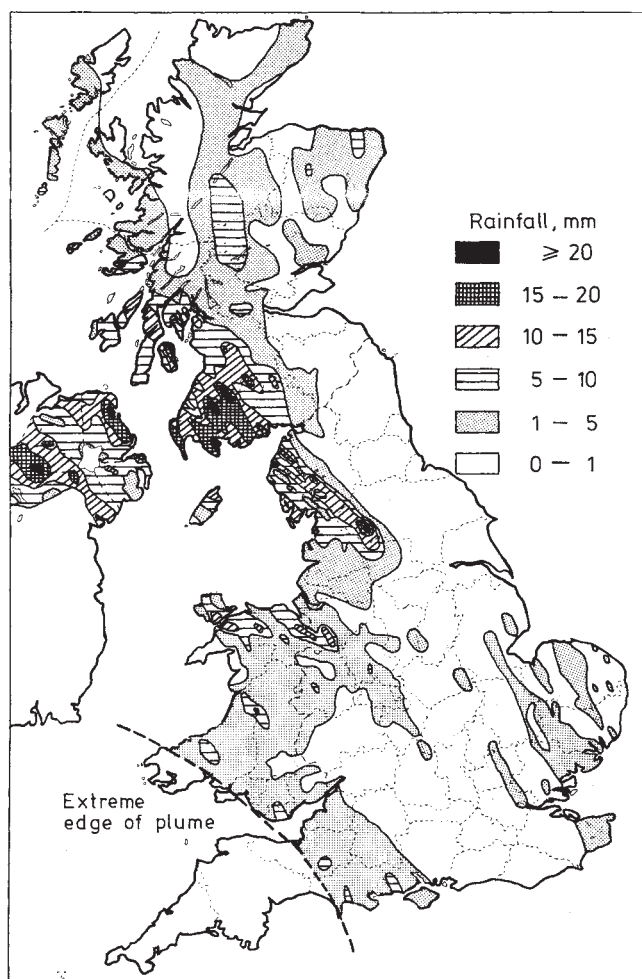


Fig. 1 Rainfall in the UK during the passage of the Chernobyl plume, 2-4 May 1986, compiled from radar data and measurements at 4,000 rain gauge stations. This is not the total rainfall for that period, only that which intercepted the plume, based on air concentration measurements.

radiation^{6,7}. For example, the concentration of ^{131}I in prolonged rain in Cumbria was 4 kBq l^{-1} . The gamma dose-equivalent rates there were typically increased to $0.4 \mu\text{Sv h}^{-1}$ above a normal background between $0.1\text{--}0.2 \mu\text{Sv h}^{-1}$, reflecting the deposition on the ground mainly of the shorter-lived gamma emitters such as ^{131}I and ^{132}Te . In regions unaffected by rainfall, the increases in background gamma radiation were much less and levels of $0.02 \mu\text{Sv h}^{-1}$ above background were typical.

For radiological protection reasons, most of the deposition data available at an early stage in the United Kingdom was on grass^{6,7}, because these data could be related to subsequent levels in cows' milk and other foodstuffs⁸. The deposition densities on grass measured during 3-6 May 1986 are given in Table 1 as county averages⁶. In regions where there was little or no rainfall, the $^{131}\text{I}/^{137}\text{Cs}$ ratio on grass was about 18. This high ratio reflects a higher dry deposition velocity for ^{131}I than for ^{137}Cs , in combination with higher integrated air concentrations of ^{131}I (particulate and vapour). In areas where there was light rainfall, the $^{131}\text{I}/^{137}\text{Cs}$ ratio was lower by a factor of 5, and where there was heavy rainfall, the ratio decreased to below 2. For example, in southeast England there was only light rainfall and isolated thunderstorms during the early hours of 3 May, as the plume moved northwards, and the $^{131}\text{I}/^{137}\text{Cs}$ ratio on grass was about 4. In northwest England and southwest Scotland, where there was heavy rainfall during storms, the $^{131}\text{I}/^{137}\text{Cs}$ ratios observed on grass were directly comparable to the observed ratios of

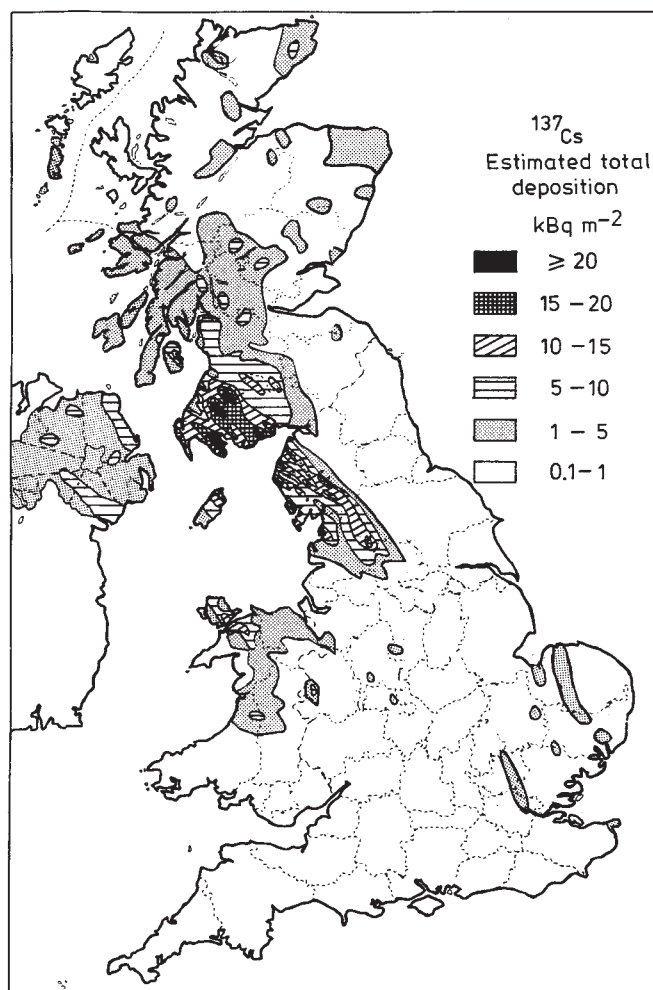


Fig. 2 Estimated total deposition of ^{137}Cs (kBq m^{-2}) over the United Kingdom due to Chernobyl releases, calculated from a washout factor of 6.5×10^5 , the rainfall data and air concentrations.

particulate species in the air, which were generally less than 2 (refs 6, 7).

The majority of data published on air concentrations during May 1986 are for membrane filters which collect only particulate material. However, data for charcoal filters from some monitoring stations indicate that the particulate fraction of ^{131}I in air was between 25 and 50%, the remainder being in the vapour phase^{6,7,9-11}. In contrast, the radiocaesium was mostly particulate and these different features have to be considered when estimating wet and dry deposition of ^{131}I and ^{137}Cs . Although dry deposition of ^{131}I and ^{137}Cs would have occurred in all parts of the country crossed by the plume, it is only in those parts that experienced little or no rain that its magnitude can be assessed from measured deposition. The observed depositions on grass in these counties is given in Table 1, and dry deposition velocities v_d (m s^{-1}) can be obtained by dividing the deposition density (Bq m^{-2}) by the measured integrated air concentrations (Bq s m^{-3}). Taking account of the total air concentration of ^{131}I , and assuming that most of the dry deposited material was retained on grass for the first few days, the retention half-period being about 5 days^{11,12}, we find that the dry deposition velocities for ^{131}I and ^{137}Cs were approximately $3 \times 10^{-3} \text{ m s}^{-1}$ and $5 \times 10^{-4} \text{ m s}^{-1}$ respectively. Table 1 shows that deposition measurements in Oxfordshire averaged 459 Bq m^{-2} of ^{131}I and 22 Bq m^{-2} of ^{137}Cs , and further measurements indicate^{6,9} a 24 h average concentration of 2 Bq m^{-3} for ^{131}I (particulate and vapour) and 0.5 Bq m^{-3} for ^{137}Cs , which give calculated values for v_d of

$2.7 \times 10^{-3} \text{ m s}^{-1}$ and $5.1 \times 10^{-4} \text{ m s}^{-1}$ for ^{131}I and ^{137}Cs respectively. These values are average deposition velocities for a diurnal cycle, and are consistent with Cambray's estimates¹¹ and those made by Chamberlain after the Windscale accident in 1957¹³.

Turning now to wet deposition effects, we can estimate the rainfall during 2-4 May 1986 that intercepted the Chernobyl plume, using surface gauge and radar data from the Meteorological Office, the air concentration data¹⁴ and trajectory analysis. As shown in Fig. 1, rainfall varied from 20 mm or more in parts of northwest England and southwest Scotland, to nil in parts of southern and eastern England. Wet deposition W (Bq m^{-2}) on a surface is assumed to be proportional to the product of rainfall R (m) and average concentration in air C (Bq m^{-3}) during the period of rainfall. Thus $W = w_r CR$, where w_r is the dimensionless washout factor defined¹⁵ as the ratio of the concentration in surface-level precipitation to the concentration in surface-level air. An interception factor f_i can be introduced into the equation to allow for the differing abilities of surfaces to intercept washout material. For soil, f_i can be assumed to be 1, but for grass the factor can vary from 0.05 to 0.3 (refs 16-18). An analysis of the available monitoring data indicates that the product $w_r f_i$ for deposition of ^{131}I and ^{137}Cs

Table 1 County averages for deposition densities on grass⁶ during 3-6 May 1986

County or region	Deposition density on grass (Bq m ⁻²)		No. of samples (n)	¹³¹ I/ ¹³⁷ Cs ratio (r)	Average ratio (n × ¹³¹ I)/ (n × ¹³⁷ Cs)
	¹³¹ I	¹³⁷ Cs			
<i>Predominantly dry areas</i>					
Berkshire	242	20	5	12	18
Channel Islands	151	4	4	38	
Cleveland	70	17	2	4	
Oxfordshire	459	22	14	21	
<i>Some light rainfall</i>					
Cheshire	325	93	6	3.5	3.8
Dorset	64	25	2	2.6	
Essex	1,345	285	1	4.7	
Gloucestershire	1,356	351	2	2.9	
Kent	2,346	—	1	—	
Lancashire	1,327	358	8	3.7	
<i>Predominantly wet areas</i>					
Cumbria	2,246	1,334	42	1.7	1.8
Dumfries and Gal- loway	610	598	14	1.0	
Gwynedd	1,009	1,235	2	0.8	
Highlands	3,604	1,164	7	3.1	
Orkney	3,159	682	4	4.6	
Strathclyde	3,445	2,285	2	1.5	
West Yorkshire	1,259	680	4	1.9	
North Yorkshire	801	470	10	1.7	

on grass is of the order of 10^5 , if we use particulate air concentrations and rainfall data for 2-4 May 1986. For example, in Cumbria the average rainfall during the passage of the plume was 10 mm, and assuming average particulate air concentration of 1 Bq m^{-3} for ^{131}I and 0.5 Bq m^{-3} for ^{137}Cs during rainfall, we find that the observed average deposition levels on grass in Table 1 of 2.2 kBq m^{-2} and 1.3 kBq m^{-2} imply that $w_r f_i$ was over 2×10^5 . If the same assumptions are made for Dumfries and Galloway, the observed average deposition levels on grass implies values for $w_r f_i$ less than 1×10^5 . These calculations show the potential uncertainties in using aggregated grass monitoring data to estimate deposition. The data came from a variety of laboratories, and no allowance has been made for different types of grass or for other variable factors. For example, although the grass measurements in Table 1 were made soon after the passage of the plume, some losses will have occurred^{11,12}. Despite these caveats, early grass measurements avoid the need to correct for weapons fallout, and variations due to surface water run off and drainage are likely to be small, compared with later soil measurements. We therefore assume a value of 1.5×10^5 for $w_r f_i$ on grass, but recognizing that f_i must vary very significantly, this value must have an associated uncertainty of at least a factor of 2.

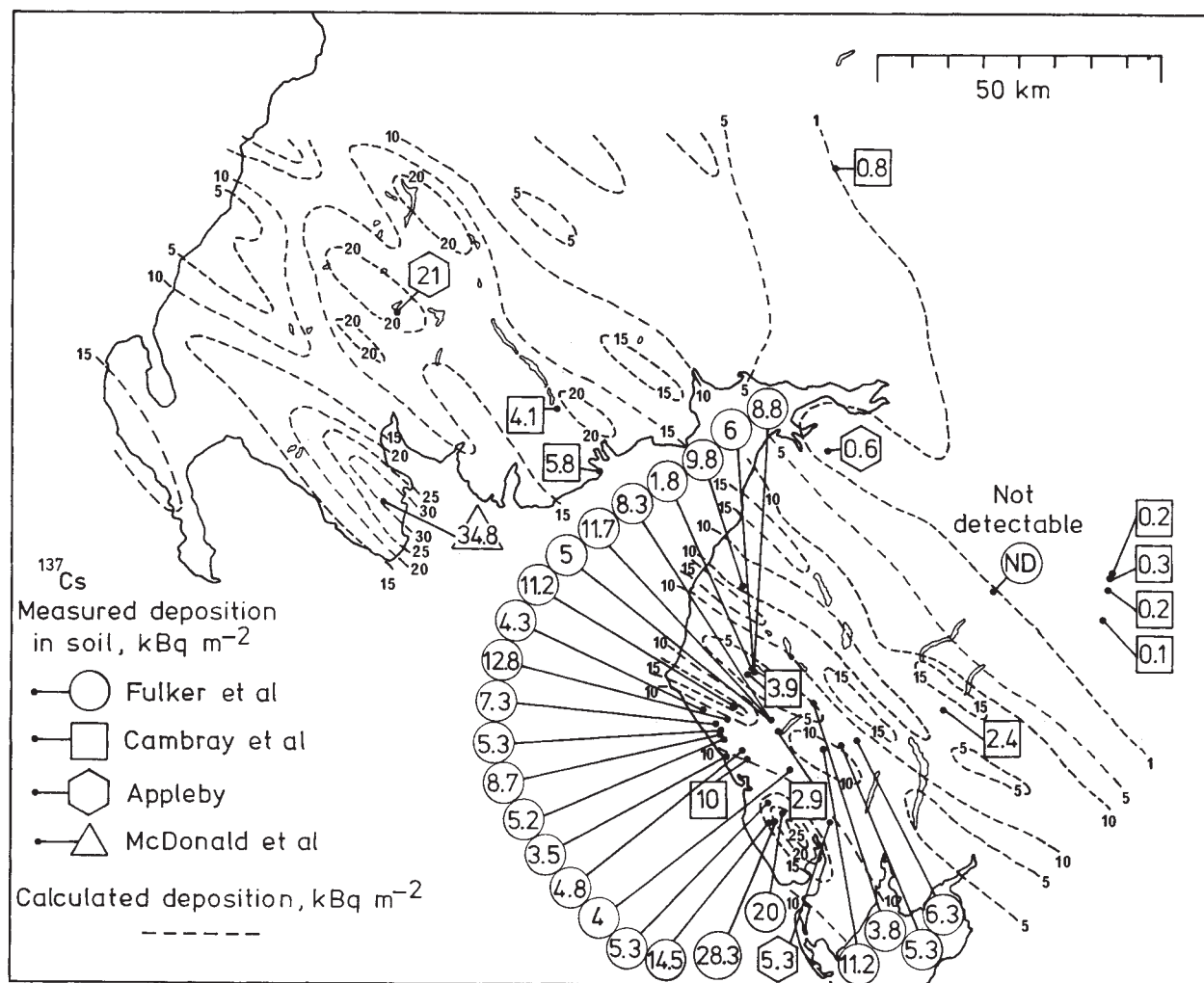


Fig. 3 Estimated total deposition of ^{137}Cs (kBq m^{-2}) in SW Scotland and Cumbria from Chernobyl, along with measurements of ^{137}Cs in soil. All measurements have been corrected for ^{137}Cs from weapons' fallout, assuming a $^{134}\text{Cs}/^{137}\text{Cs}$ ratio of 0.6 for Chernobyl fallout. Measurements of ^{137}Cs in soil kindly provided by R. S. Cambray (Harwell), M. J. Fulker (BNFL Sellafield), P. Appleby (Univ. Liverpool) and P. McDonald and M. S. Baxter (Scottish Univs. Research and Reactor Centre).

At Holmrook in west Cumbria, Cambray¹¹ reported that the total ^{137}Cs in the soil was 10 kBq m^{-2} , discounting previous weapons fallout. Meteorological data shows that two storms affected Holmrook during the passage of the plume, the first giving 5 mm of rain when the air concentration of ^{137}Cs was elevated to 3 Bq m^{-3} and the second gave 2 mm of rain when the air concentration was still somewhat above average at 1 Bq m^{-3} . These data yield a value of 5.9×10^5 for w_r in the case of ^{137}Cs , and similar analyses at other locations where measurements are available^{10,11,18} give an average close to 6.5×10^5 . This implies that f_i was 0.23, which is close to the standard interception factor of 0.25 assumed for grass^{16,17}, based on a wide range of experiments with particulate species. Also the derived washout factor can be compared with a value of $(1 \pm 0.3) \times 10^6$ for the average ($\pm$ one standard deviation) of measurements of both natural and bomb debris radionuclides¹⁵.

Using the washout factor of 6.5×10^5 and interpolated concentrations in air, we can estimate the total deposition of ^{137}Cs at each of over 4,000 rain-gauge stations throughout the United Kingdom, and these values have, in turn, been interpolated to give the contours shown in Fig. 2. Radar measurements at the time showed narrow bands of rain moving across the country and, where they intercepted the Chernobyl plume, radionuclide deposition would also have occurred in narrow bands. More detailed contours for southwest Scotland and Cumbria are given in Fig. 3, along with measurements of ^{137}Cs in soil. There is

good general agreement, although some caution is required when interpreting the results. Rain gauge stations are normally in accessible places, below the highest land, and can therefore underestimate rainfall in mountainous areas. Furthermore, measurements of deposition are not normally coincident with rain gauge stations, and rainfall can vary over short distances. The amount of ^{137}Cs in soil also varies over short distances, as exemplified by Fulker's measurements in Cumbria^{10,18}, where results varied from 1.8 to 8.8 kBq m^{-2} at one location, and from 5 to 11.7 kBq m^{-2} at another. Variations in altitude, drainage, soil type and land usage contribute to this.

It is possible to estimate the total amount of ^{137}Cs and ^{131}I deposited on the United Kingdom from Chernobyl, using the dry deposition velocities and washout factors derived above. The average integrated particulate air concentration for ^{137}Cs was about $20 \text{ Bq m}^{-3} \text{ h}$ in the United Kingdom⁶. This gives an average dry deposition of 36 Bq m^{-2} if we use a dry deposition velocity of $5 \times 10^{-4} \text{ m s}^{-1}$. The average total wet deposition on the ground is estimated at $1,300 \text{ Bq m}^{-2}$ if we use the washout factor of 6.5×10^5 and assume an average air concentration of 1 Bq m^{-3} during rainfall with an average rainfall of 2 mm for the United Kingdom. The average total deposition was therefore about 1.3 kBq m^{-2} , and the total deposition of ^{137}Cs on the United Kingdom (area $2.5 \times 10^{11} \text{ m}^2$) would have been about $3 \times 10^{14} \text{ Bq}$. These estimates are supported by a simple integration of the deposition values given in Fig. 2 which yields

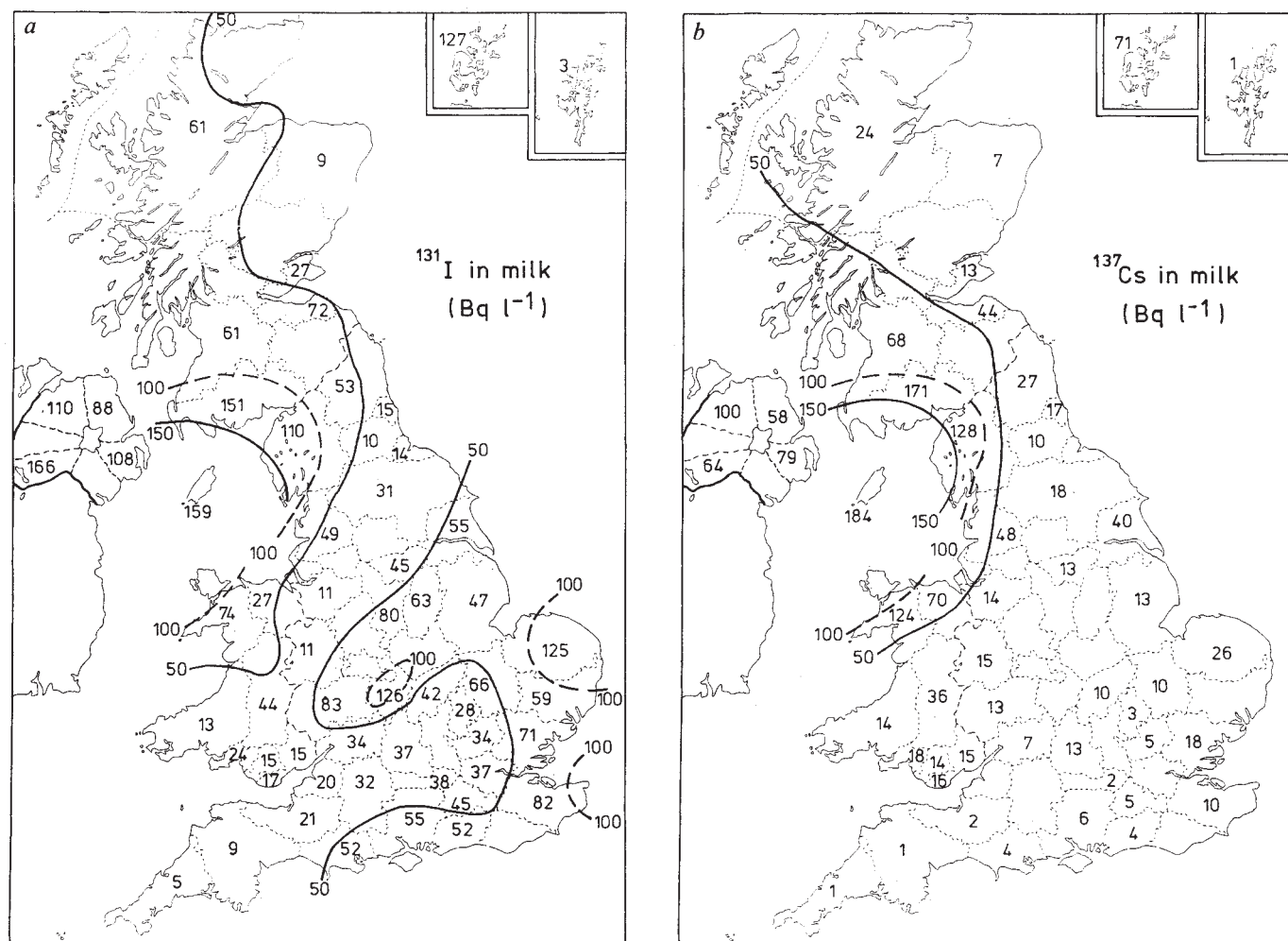


Fig. 4 a, Average concentrations of ^{131}I in cows' milk (Bq l^{-1}), 5–8 May 1986, from monitoring data⁶. b, Average concentrations of ^{137}Cs in cows' milk (Bq l^{-1}), 5–8 May 1986, from monitoring data⁶.

1.2 kBq m^{-2} , and are in agreement with Cambray's estimate of $3.2 \times 10^{14} \text{ Bq}$ for total deposition in the UK¹¹.

The total deposition of ^{131}I from Chernobyl was considerably more complicated than for ^{137}Cs , because of the relatively high dry deposition velocity of ^{131}I and the presence of both vapour phase and particulate iodine species in the plume. Reported values for ^{131}I in the top 5 cm of soil, including grass, show ^{131}I to ^{137}Cs ratios of about 5:1 in the predominantly wet areas (ref. 11 and Fig. 3). This is confirmed by measured ratios in rain, although the scatter about an average 5:1 ratio is greater. In contrast to the levels observed on grass, the total deposition reflects the additional contribution of gaseous ^{131}I which was mainly washed down into the soil. These observations imply that the total depositions of ^{131}I on the United Kingdom must have been around $1.5 \times 10^{15} \text{ Bq}$, and allowing for its high dry deposition velocity relative to ^{137}Cs , we estimate the total deposition might have approached $2 \times 10^{15} \text{ Bq}$.

As a percentage of the total releases from the Chernobyl accident, the amounts of ^{131}I and ^{137}Cs deposited on the United Kingdom were about 0.7% and 0.8% respectively, based on Soviet estimates of $3 \times 10^{17} \text{ Bq}$ and $4 \times 10^{16} \text{ Bq}$ for the respective total releases¹. These percentages are uncertain by at least a factor of two, and if Cambray's estimate of $5 \times 10^{16} \text{ Bq}$ for the total ^{137}Cs release is used¹¹, the percentage of ^{137}Cs deposited on the United Kingdom becomes 0.6%. An analysis of the trajectories of the plume indicates that the portion of the release affecting the United Kingdom left the site on Saturday 26 April^{2,3} and the Soviet authorities estimate that, on that first day, about $1 \times 10^{17} \text{ Bq}$ of ^{131}I and $1 \times 10^{16} \text{ Bq}$ of ^{137}Cs were released^{1,4}. The corresponding percentage of this release affecting the United

Kingdom was approximately 2% for ^{131}I and 3% for ^{137}Cs , if we use the average total depositions derived above. The fraction of ^{131}I deposited on the United Kingdom was less than that for ^{137}Cs because of its short half life (8.05 days) and its higher dry deposition en route.

The deposition of ^{131}I and ^{137}Cs led directly to incorporation in the food chain, and average levels observed in cows' milk in UK counties during 5–8 May, are shown in Fig. 4^{6,8}. The levels of ^{131}I in milk in parts of southern England were quite comparable to levels observed in Cumbria and other high rainfall areas of the northwest, but were not accompanied by similar levels of ^{137}Cs . The levels in milk should reflect deposition on pasture grass, and Table 1 also shows that levels of ^{131}I on grass in Kent and Essex were comparable with the northwest, whereas ^{137}Cs levels were low. Wilkins *et al.*¹⁹ have shown that transfer factors in the milk pathway vary in different parts of the country, depending on agricultural practices. However, monitoring data for other foodstuffs and in environmental materials show similarly elevated ^{131}I levels in parts of the south, and so dry deposition seems to be the most likely cause here. The levels of ^{131}I on grass in the areas which experienced little or no rain were of the order of 2 kBq m^{-2} or less, most of which was dry deposited and retained on grass. In areas where there was substantial rainfall, both during and after the passage of the Chernobyl plume, total ^{131}I deposition was much higher, but a smaller fraction appears to have been retained on grass. For example, measurements of total ^{131}I deposition in the top 5 cm of soil have been reported in the range $10\text{--}40 \text{ kBq m}^{-2}$ in Cumbria^{10,11}, where measurements on grass were typically 4 kBq m^{-2} or less^{6,8}. In contrast, ^{137}Cs deposition in dry areas was low

because of its low dry deposition velocity, but there was significant wet deposition in the higher rainfall areas. The difference between observed levels of ^{137}Cs on grass, in foodstuffs and other environmental materials in wet and dry areas, is therefore much more evident, as demonstrated by the cows' milk data shown in Figure 4b. These observations confirm that wet deposition during heavy rainfall is an important mechanism for both ^{131}I and ^{137}Cs , but dry deposition of ^{131}I can also contribute significantly to levels in the food chain.

We thank the many colleagues who obtained and compiled the monitoring data and Mrs Jean Fisher (NRPB) who prepared the figures. We are grateful to Dr Peter Appleby (University of Liverpool), Roger Cambray (Harwell Laboratories), Dr Mike Fulker (BNF Sellafield), Mr P McDonald and Professor Baxter (Scottish Universities Research and Reactor Centre) for providing details of their measurements of ^{137}Cs in soil.

Received 4 August 1987; accepted 14 January 1988.

1. *Vienna Safety Series 75-INSAG-1* (International Atomic Energy Agency, Vienna, 1986).
2. Smith, F. B. & Clark, M. J. *Nature* **322**, 690-691 (1986).
3. ApSimon, H. M., MacDonald, H. F. & Wilson, J. J. *J. Soc. Radiol. Prot.* **6**, 109-119 (1986).
4. DOE/ER-0332 (US Department of Energy, Washington, 1987).
5. Morrey, M. *et al.* V/E/1 Contract 86 398 (CEC Luxembourg, 1987).
6. *Levels of Radioactivity in the UK from the Accident at Chernobyl* (HMSO, London, 1986).
7. HS/R/206/86 Issue A (Central Electricity Generating Board, London, 1986).
8. *Post-Chernobyl Monitoring in England and Wales* (HMSO, London, 1987).
9. Fry, F. A., Clarke, R. H. & O'Riordan, M. C. *Nature* **321**, 193-195 (1986).
10. Jackson, D., Jones, S. R., Fulker, M. J. & Coverdale, N. G. *M. J. Soc. Radiol. Prot.* **7**, 75-87 (1987).
11. Cambray, R. S. *et al.* *Nucl. Energy* **26**(4), 77-101 (1987).
12. Fry, F. A. *Br. J. Radiol.* **60**, 1147-1158 (1987).
13. Chamberlain, A. C. *Int. J. Air Pollut.* **3**, 63-88 (1960).
14. Clark, M. J. & Smith, F. B. NRPB-R (in the press).
15. Slinn, W. G. N. in *Atmospheric Science and Power Production* (ed. Randerson, D.), 466-532 (US Department of Energy, Washington, 1984).
16. Chamberlain, A. C. *Atmos. Environ.* **4**, 57-78 (1970).
17. Simmonds, J. R., Linsley, G. S. & Jones, J. A. NRPB-R89 (HMSO, London, 1979).
18. Fulker, M. J. *J. Environ. Radioactivity* **5**, 235-244 (1987).
19. Wilkins, B. T., Bradley, E. J. & Fulker, M. J. *Sci. Tot. Env.* (in the press).

Two new three-dimensional twelve-ring zeolite frameworks of which zeolite beta is a disordered intergrowth

M. M. J. Treacy & J. M. Newsam

Exxon Research & Engineering Company, Annandale,
New Jersey 08801, USA

Zeolite beta, first described in 1967¹, is an active catalyst and a useful sorbent¹. Sorption^{1,2} and catalytic data^{3,4} suggest that the zeolite could possess a three-dimensional 12-ring pore system. Such a pore system suggests technological potential similar to that of faujasite framework materials, but until now the structure of this zeolite has eluded determination. Powder X-ray diffraction patterns comprise both sharp and broad features, indicative of an extensively faulted structure. Here we determine the structure of zeolite beta by high-resolution electron microscopy, electron diffraction and computer-assisted modelling. The zeolite is an intergrown hybrid of two distinct but closely related structures. Both are constructed from the same centrosymmetrical tertiary building unit arranged in layers, and both possess three-dimensional 12-ring pore systems. One end member, polymorph A, forms an enantiomorphous pair, with symmetries $P4_222$ and $P4_222$, with $a = 12.4 \text{ \AA}$ and $c = 26.5 \text{ \AA}$. Polymorph B, in which the stacking of layers alternates in handedness, is achiral with space group $P\bar{1}$, and $a \approx b = 12.4 \text{ \AA}$, $c = 14.5 \text{ \AA}$, $\alpha \approx \beta = 73^\circ$, $\gamma \approx 90^\circ$. The high density of stacking faults in zeolite beta materials arises because successive layers must interconnect in either a left- or a right-handed fashion, and both modes of linkage occur with almost equal probability.

Zeolite beta samples were prepared from $\text{Na}_2\text{O} \cdot \text{TEA}_2\text{O} \cdot \text{Al}_2\text{O}_3 \cdot \text{SiO}_2 \cdot \text{H}_2\text{O}$ gels (TEA, tetraethylammonium cation)^{1,5}. Bright-

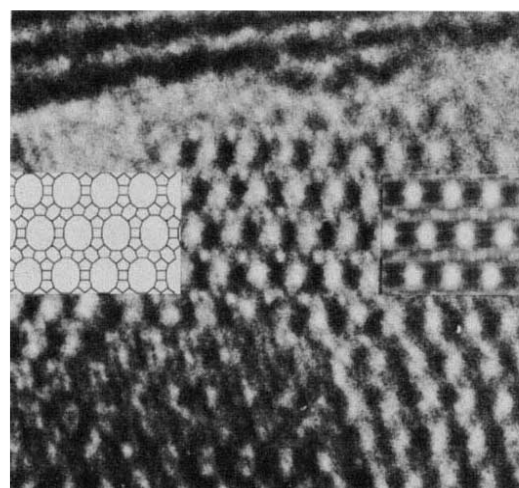
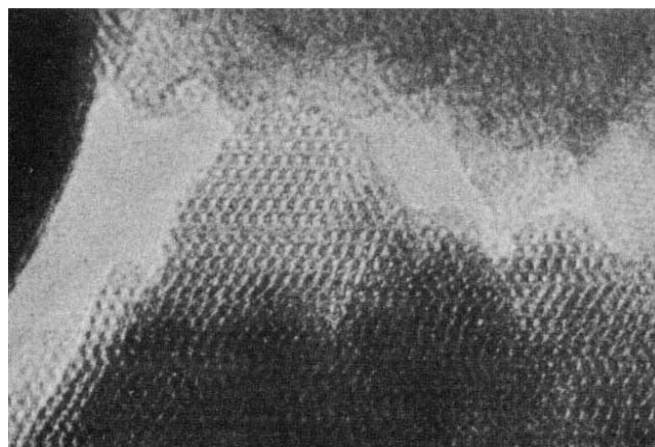


Fig. 1 *a*, Bright-field micrograph of a zeolite beta crystallite viewed perpendicularly to the 'square' (c) axis. The plane stacking along the 'square' axis direction is highly faulted (see text). *b*, An expanded local section of a similar micrograph which suggests the secondary ring structure for this projection shown in the left inset (at the resolution of these micrographs ($\sim 3.5 \text{ \AA}$), structure images can be interpreted in terms of the projected potential of the structure). The computed image¹¹ based on the derived structure of polymorph A in this projection is given as the right inset.

field electron micrographs show that crystallites are ~ 0.1 – $0.6 \text{ \mu m}$ in diameter, with square profiles when viewed along one axis. Perpendicularly, crystallites show irregular obliquely inclined pyramidal features whose peaks point along the 'square' axis. A high-resolution micrograph taken perpendicularly to the 'square' axis and parallel to one of the square sides (Fig. 1) shows these pyramidal features and the projection of the pore structure (large white features), showing considerable stacking disorder. Pore stacking can be roughly categorized into two sequences, ABAB... type and ABC... type, because pores in successive layers are shifted laterally by $\pm 1/3$ of an intralayer pore spacing. Sorption^{1,2} and catalytic^{3,4} properties indicate that the main pore structure is accessed through 12-rings. Ion-exchange and sorption data indicate that the largest secondary cage windows are, at most, 6-rings, and the synthesis conditions, zeolite composition and infrared data imply that the zeolite contains 5-rings. Structure images such as Fig. 1 show that the intergrowths arise from twinning at planes that bisect the main pores. The framework topology at the fault plane must therefore support what appears in projection to be a mirror operation. Furthermore, such structure images provide clues as to the

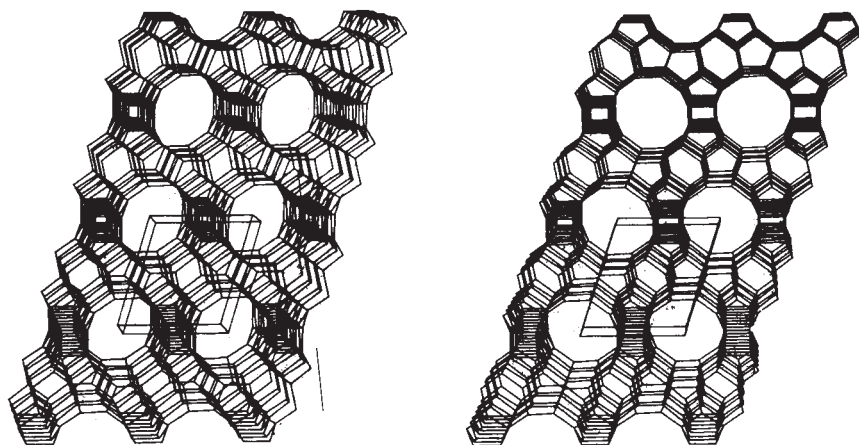


Fig. 2 Stereoview of the framework structure of the ABC... polymorph B which is achiral. The framework is represented as straight lines connecting adjacent T-sites. The oxygen atoms (not shown) lie close to the midpoints of these lines.

secondary cage structure around the pore apertures (in the projection approximation, 6-rings are more pronounced than 5-rings which, in turn, are more pronounced than 4-rings). The projection of the structure in this direction that was derived from such lattice images is shown in Fig. 1. This projection resembles the [010] projection of ZSM-12 (ref. 6) (MTW-framework⁷), although expanded by ~12% in the interlayer and by 2% in the intralayer directions.

Information about the third crystallographic direction and possible space-group symmetries was derived from a series of electron-diffraction experiments. Views down the unique square axis, c , possess accurately 4-fold symmetry (averaged through a large number of fault planes), implying that the perpendicular axes, a and b , are mutually orthogonal and have equal lengths. Diffraction patterns taken obliquely, together with extensive tilting and survey experiments, show that views down both the a and b axes (as averaged by the twinning operation) are similar. Diffraction patterns containing the c^* axis show streaking parallel to c^* when h or $k \neq 3n$. Streak widths are consistent with a fault probability of $40(\pm 10)\%$ in the ABC... stacking sequence, in good agreement with similar estimates derived from images. Precisely aligned patterns show a repeat distance of $26.5 \text{ \AA}$ along c (twice the interlayer spacing observed in the images) and show $00l$, $l \neq 4n$ extinctions. It follows, therefore, that the operation relating a pore centre viewed down the a axis with that seen down b is a $c/4$ shift ($\sim 6.6 \text{ \AA}$ along the c axis) combined with a $\pm 90^\circ$ rotation. The ABAB... sequence (termed polymorph A), with successive $c/4 + 90^\circ$ (or successive $c/4 - 90^\circ$) stacking vectors, realizes a 4_1 (or 4_3) screw axis. The absence of other systematic extinctions restricts the possible space groups to $P4_1$ (or $P4_3$) or $P4_122$ (or $P4_322$). The projection derived from the lattice-imaging experiments (Fig. 1) is consistent with the presence of 2-fold axes perpendicular to c , and therefore $P4_122$ was initially chosen. Computer-assisted modelling was used to derive a sensible three-dimensional T-atom connectivity (T: Si, Al) from the projection shown in Fig. 1 and the symmetry operations of space group $P4_122$. Oxygen atoms were added at the mid-points of the T-T vectors, and the full structure then optimized by distance least squares (d.l.s.) (refs 8-10). The d.l.s.-optimized lattice constants are $a = 12.66 \text{ \AA}$ and $c = 26.41 \text{ \AA}$. Least-squares optimized lattice constants of $a = 12.447(5) \text{ \AA}$ and $c = 26.56(1) \text{ \AA}$ were derived from the powder X-ray diffraction (PXD) pattern. Approximate T-atom coordinates for polymorph B, with the ABC... sequence (in which layers stack alternately $c/4 + 90^\circ$ and $c/4 - 90^\circ$), were obtained by transforming the corresponding coordinates of the polymorph A structure, and the framework geometry optimized similarly (data available on request from authors before publication). Polymorph B is described by a triclinic cell, $P\bar{1}$, with d.l.s.-optimized lattice constants of $a = 12.69 \text{ \AA}$, $b = 12.70 \text{ \AA}$, $c_T = 14.42 \text{ \AA}$, $\alpha = 73.2^\circ$, $\beta = 72.9^\circ$ and $\gamma = 90.1^\circ$. Corresponding

optimized values of $a = 12.48(6) \text{ \AA}$, $b = 12.50(1) \text{ \AA}$, $c_T = 14.84(2) \text{ \AA}$, $\alpha = 73(1)^\circ$, $\beta = 70(1)^\circ$ and $\gamma = 87(1)^\circ$ were obtained from analysis of the PXD pattern. A stereoview of the structure of polymorph B is shown in Fig. 2. The results of a multislice lattice image calculation¹⁰ based on the polymorph A structure are included in Fig. 1. The distribution of intensities and peak widths observed in the PXD patterns of zeolite beta materials are reproduced qualitatively by a combination of the computed patterns for the two polymorphs, assuming an intergrowth probability of ~50%. Full quantitative descriptions of both unfaulted and faulted PXD patterns will be given elsewhere (J.M.N., M.M.J.T., W.T. Koetsier and C. B. deGruyter, manuscript in preparation).

In common only with the faujasite framework⁷, both structures have fully three-dimensional pore systems with 12-rings as the minimum constricting apertures. The 12-ring channels along a

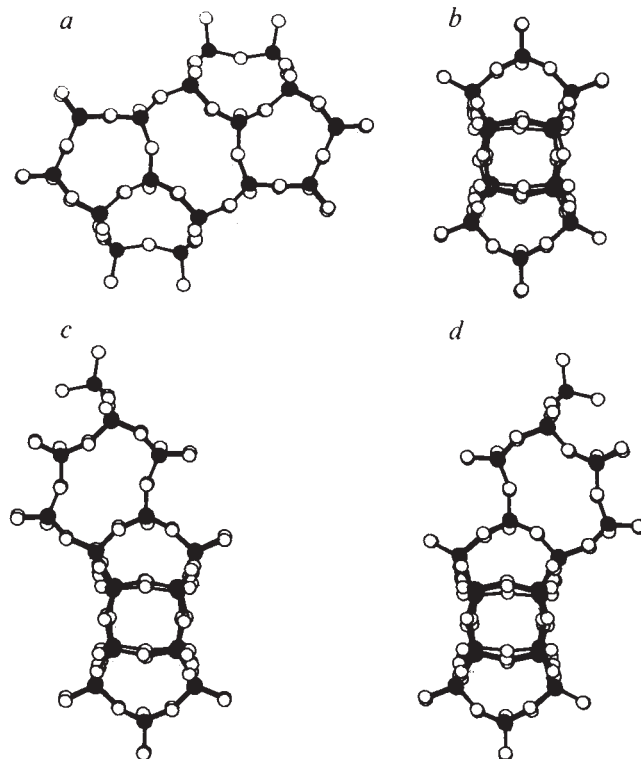


Fig. 3 *a*, The basic structural building block with approximate $2/m$ symmetry. *b*, Same unit viewed after a 90° rotation about the vertical axis. A second, perpendicular 6-ring can be added to create an object with c , left-handed or *d*, right-handed symmetry. (T atoms, solid; O atoms, open circles; fragments were clipped from the polymorph B structure illustrated in Fig. 2.)

and *b* are similar in both structures, but the pore system along *c* in polymorph A is somewhat more tortuous than in polymorph B. The common stacking faults affect only this net tortuosity along *c*, but do not significantly affect the accessible pore volume. The tertiary building block of these structures has no intrinsic preference for right- or left-handed connectivity (Fig. 3), explaining the observed nearly random extent of stacking faults. The high concentrations of framework hydroxyl groups that many zeolite beta materials support probably arise from unsatisfied linkages introduced by intralayer left-/right-handedness indecisions. The two structures described here represent the two end-members of a family of structures of which beta zeolites comprise only a small subset (that represented by close to maximal intergrowth probability). Another interesting member of this family has regular ABAB... stacking when viewed along *a*, but regular ABC... stacking when viewed along *b*. Other regular structures are also possible, although apparently none has yet been accessed synthetically.

We thank colleagues for samples, suggestions and helpful discussions, particularly R. Hulme, H. E. Robson, W. T. Koetsier, A. J. Jacobson, B. Flannery and C. B. deGruyter.

Received 17 November 1987; accepted 11 January 1988.

1. Wadlinger, R. L., Kerr, G. T. & Rosinski, E. J. *US Patent No. 3,308,069* (Mobil Oil Corporation, 1967).
2. Bond, A. E., Burgess, C. G. V. & Martin, D. E. *US Patent No. 3,793,385* (BP PLC, 1974).
3. Martens, J. A., Tielens, M., Jacobs, P. A. & Weitkamp, J. *Zeolites* 4, 98-107 (1984).
4. Martens, J. A., Perez-Pariente, J. & Jacobs, P. A. *Acta phys. chem.* 31, 487-495 (1985).
5. Perez-Pariente, J., Martens, J. A. & Jacobs, P. A. *Appl. Catal.* 31, 35-64 (1987).
6. LaPierre, R. B. *et al. Zeolites* 5, 346-348 (1985).
7. Meier, W. M. & Olson, D. H. *Atlas of Zeolite Structure Types* 2nd edn (Butterworth, Oxford, in the press).
8. Meier, W. M. & Villiger, H. Z. *Kristallogr. Kristallgeom.* 129, 411-423 (1969).
9. Baerlocher, C., Hepp, A. & Meier, W. M. *DLS-76 Manual* (ETH, Zurich, 1977).
10. O'Keefe, M. A. & Buseck, P. R. *Trans. Am. crystallogr. Ass.* 15, 27 (1979).

Evidence from X-ray absorption for network-forming Fe^{2+} in molten alkali silicates

G. A. Waychunas*, G. E. Brown Jr*†, C. W. Ponader† & W. E. Jackson†

* Center for Materials Research and † Department of Geology, Stanford University, Stanford, California 94305, USA

Fe^{2+} is the most abundant iron species in magmas and in many slags. Its structural role in these liquids is poorly understood, largely because of the difficulty of studying melts at high temperatures by direct structural methods. Whether Fe^{2+} behaves as a network modifier, network former or free-ion complex has not been adequately resolved, yet this is fundamental to understanding the properties of silicate liquids. Also uncharacterized are the structural changes accompanying the melt-to-glass transition in Fe-bearing silicate melts. Here we report the results of a high-temperature synchrotron-based X-ray absorption study of Fe in silicate glasses and melts of compositions near $\text{Na}_2\text{FeSi}_3\text{O}_8$ and $\text{K}_2\text{FeSi}_3\text{O}_8$. The glasses were also analysed by ^{57}Fe Mössbauer spectroscopy. We conclude that Fe^{2+} is a four-coordinated network former in these melt/glass systems and that little structural relaxation occurs at the iron site during the melt-to-glass transition. These results and structural data for $\text{Fe}_2^{2+}\text{SiO}_4$ melts¹ suggest the possibility of a pressure-induced change from four- to six-coordination for Fe^{2+} in magmas in the Earth's upper mantle.

The samples were prepared from mixtures of oxides and carbonates by melting at $\sim 400^\circ\text{C}$ above liquidus temperatures in iron crucibles under an argon atmosphere. Optical examination revealed small blebs of Fe metal in an otherwise

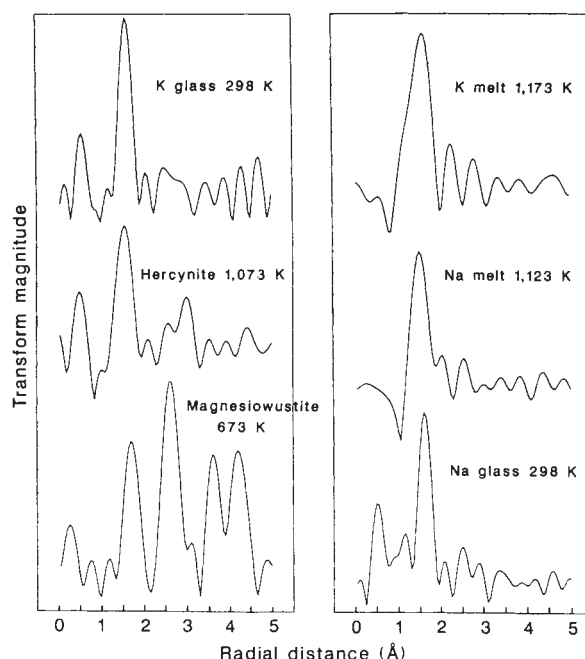


Fig. 1 Fourier transforms of the weighted EXAFS functions of alkali-iron glasses and melts compared with model compounds having mainly Fe^{2+} tetrahedral (hercynite) and octahedral (magnesioiwüstite) coordination. Least-squares fitting of the first-shell back-Fourier transform yields the following values for coordination number (CN), first-neighbour mean bond length (*R*) and refined Debye-Waller Factor (σ^2); the values in parentheses are the one-standard-deviation uncertainties. Magnesioiwüstite at 673 K: CN=6.21 (0.62), *R*=2.144 (0.006), σ^2 =0.0077 (0.0014); hercynite at 1,073 K: CN=4.02 (0.59), *R*=1.965 (0.006), σ^2 =0.0059 (0.0016); K glass at 298 K: CN=4.05 (0.49), *R*=1.999 (0.004), σ^2 =0.0010 (0.0017); Na glass at 298 K: CN=3.29 (0.40), *R*=2.018 (0.004), σ^2 =0.0018 (0.0012); K melt at 1,173 K: CN=5.01 (0.97), *R*=1.937 (0.008), σ^2 =0.0118 (0.0032); Na melt at 1,123 K: CN=3.19 (0.57), *R*=1.937 (0.008), σ^2 =0.0068 (0.0023). Phase functions used in the analysis are those calculated by Teo and Lee.¹⁷

homogeneous, crystal-free glass phase. These Fe blebs were magnetically separated before X-ray studies and avoided during microprobe analysis. Our X-ray absorption spectra show no evidence of metallic iron, which has a very distinctive spectral signature. The microprobe and Mössbauer results gave stoichiometries of $\text{Na}_{0.03}\text{K}_{1.71}\text{Fe}_{0.74}^{2+}\text{Fe}_{0.06}^{3+}\text{Si}_{3.16}\text{O}_8$ and $\text{Na}_{1.88}\text{K}_{0.01}\text{Fe}_{0.84}^{2+}\text{Fe}_{0.04}^{3+}\text{Si}_{3.08}\text{O}_8$ for the K- and Na-rich samples, respectively. Variations from the desired stoichiometry are due to alkali loss during heating and phase separation of Fe metal. Analysis of the K-rich sample after the EXAFS (extended X-ray absorption fine structure) study indicated no change in composition from the initial material.

The apparatus used for the X-ray absorption measurements consisted of a vacuum furnace with tantalum windings, with a boron nitride sample boat mated to an ionization chamber X-ray detector². The oxygen partial pressure was adjusted by regulation of the furnace vacuum. The best results required a vacuum that allowed slight oxidation of the melts; reducing conditions quickly disproportionated the Fe in the melt, producing Fe metal.

The structure functions derived from Fourier transforms of the EXAFS spectra are shown in Fig. 1. Comparison with the structure functions of hercynite and magnesioiwüstite demonstrates the similarity of Fe-O bond distances in the melts and glasses with those of hercynite and the disparity with the octahedral Fe-O distances in magnesioiwüstite. The structure functions of the melts are broadened relative to those of the glasses owing

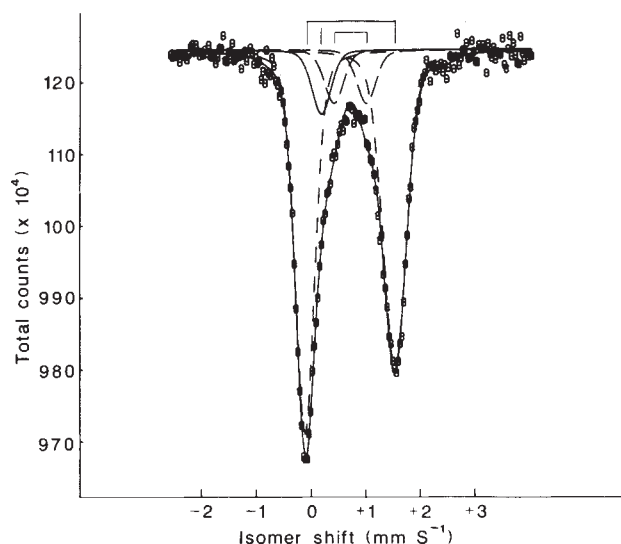


Fig. 2 Mössbauer spectrum of the sodium glass sample after original synthesis. The isomer shift is measured relative to metallic iron. Dashed lines, fitting of two Fe^{2+} doublets; solid line, singlet fit of Fe^{3+} ; circles, observed data points. The spectra of all samples as synthesized and after quench in the experimental apparatus are very similar, except for additional Fe^{3+} in the quenched samples. The small peaks consist of one of the Fe^{2+} doublets and the Fe^{3+} singlet. The lines above the fitted peaks indicate the positions of the two peaks belonging to each Fe^{2+} doublet and the Fe^{3+} singlet peak.

to the loss of higher-wavevector data at the highest temperatures of the study. The effective observation range in wavevector is $3.5\text{--}10.5 \text{ \AA}^{-1}$ at $1,200 \text{ K}$, as compared with $3.5\text{--}15 \text{ \AA}^{-1}$ at 300 K . The observed EXAFS attenuation due to thermal disorder effects was not as large as anticipated for melts, suggesting that $\text{Fe}^{2+}\text{--O}$ thermal vibrations in these melts are no larger than in silicate crystals at similar temperatures. $\text{Fe}^{2+}\text{--O}$ bond distances obtained from curve fitting of the weighted EXAFS chi functions³ are consistent with $\text{Fe}^{2+}\text{--O}$ distances found in Fe^{2+}O_4 tetrahedra in the model compounds hercynite, chromite and staurolite.

The $\text{Fe}\text{--O}$ distances, coordination numbers and Debye-Waller factors for the glasses, melts and model compounds are given in the Fig. 1 legend. These values indicate that the Fe^{2+} environment in the sodium glass and melt has greater static disorder, or range of bond lengths, than in the potassium analogue. The refined Debye-Waller factors (σ^2) in the glass samples increase with increasing temperature in a trend consistent with isotropic temperature factors, as observed in high-temperature, X-ray crystal structure refinements of Fe-bearing silicate phases⁴. However, there is no dramatic increase in σ^2 on melting of the glass samples, indicating that no significant additional static disorder occurs around the Fe site on melting in these samples. Static disorder causing highly irregular bond-length distributions can lead to EXAFS predictions of a reduced coordination number and reduced average bond length (R) in highly disordered materials. This problem should not be serious in our spectra because the observed σ^2/R values are significantly less than the predicted threshold ($0.01 \text{ \AA}$) where this problem emerges⁵. To provide confirmation of the EXAFS results, Mössbauer and XANES (X-ray absorption near-edge structure) analyses were also performed.

Mössbauer spectra of the original glass samples and quenched melts were run, and a representative example is shown in Fig. 2. In general, isomer shift (IS) values are sensitive to the average $\text{Fe}\text{--O}$ bond length⁶, but this affords no certainty of coordination number determination, as shown in Fig. 3. In this figure, T, P, O and D represent four-, five-, six- and eight-coordinated Fe^{2+} , respectively. $\text{Fe}^{2+}\text{--O}$ distances were derived from single-crystal

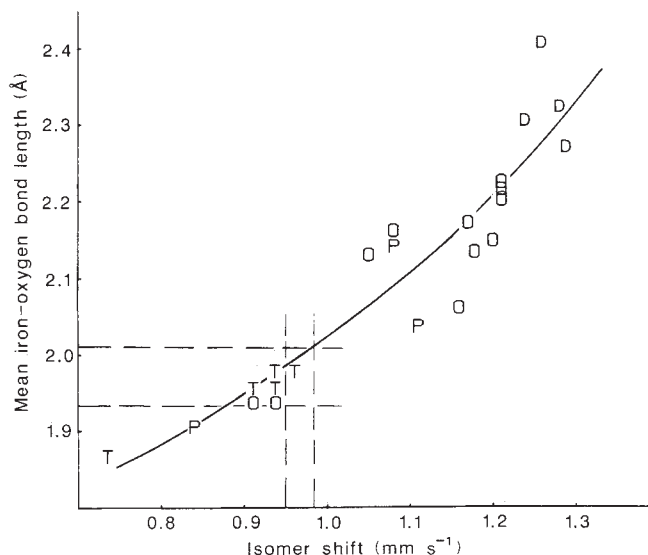


Fig. 3 Iron-oxygen bond length plotted against isomer shift (relative to metallic iron) in silicates and oxides. Dashed lines give the range of bond lengths obtained from the glasses and melts by EXAFS and the range of isomer shifts observed in the original and post-experiment glasses. Points labelled D represent eight-coordinated Fe^{2+} in spessartine, grossular, almandine and pyrope; O, six-coordinated Fe^{2+} in rhodonite (three sites), allanite, magnesio-wüstite, neptunite, fayalite (two sites), osumulite, hedenbergite and hercynite; P, five-coordinated Fe^{2+} in grandierite, vesuvianite and rhodonite; T, tetrahedral Fe^{2+} in hercynite (two separate refinements), chromite, staurolite and Ti,Fe-bearing andradite. The data come from about 40 references on crystal structure and Mössbauer spectroscopy and will be discussed in another publication. Mössbauer work on the hercynite was done in part for this paper. The d.l.s.-simulation distances were prepared for this paper as described in the text.

X-ray diffraction refinements, except in the case of low Fe^{2+} concentrations, where the $\text{Fe}^{2+}\text{--O}$ distances were estimated from distance least-squares (d.l.s.) modelling techniques⁷. The IS values in the glasses ($0.95\text{--}0.98 \text{ mm s}^{-1}$ —relative to ^{57}Fe in Fe metal) are consistent with the EXAFS-derived distances. Past EXAFS work on Fe^{2+} in crystalline fayalite has shown the effects of large static disorder on the derived distances and coordination number³, yet IS values for Fe^{2+} in fayalite are consistent with the average $\text{Fe}^{2+}\text{--O}$ distance in fayalite. This and other evidence indicates that Mössbauer IS values are largely unaffected by static disorder. Accordingly, the IS values for the glass and melt samples confirm $\text{Fe}^{2+}\text{--O}$ bond lengths characteristic of tetrahedral coordination, as in staurolite.

Additional confirming evidence for a tetrahedral site assignment for Fe^{2+} derives from the $1s \rightarrow 3d$ 'pre-edge' electronic absorption feature in the XANES spectrum of the glasses, melts and model compounds (Fig. 4). The pre-edge feature intensity is sensitive to Fe^{2+} site geometry, varying from 0.0 to 2.0% of the main edge step in octahedral sites, but increasing to 5–7% in tetrahedral sites⁸. Hercynite, with 83% of Fe^{2+} on the tetrahedral site, shows a feature with the latter intensity, as do all of the glasses and melts. Fe^{2+} in magnesio-wüstite, in a regular octahedral site, shows much lower intensity. Ferric iron, if present in tetrahedral coordination, can give rise to a pre-edge feature with up to 15% of the main edge intensity⁸, but there is far too little Fe^{3+} to account for the feature intensity in the glass samples or in hercynite.

Thus, the pre-edge feature intensity suggests either a tetrahedral site or a severely distorted higher-coordination site. To account for the pre-edge intensity, the latter would have to be more distorted than any we have yet observed. Such a distorted site might produce EXAFS results similar to ours, but not the

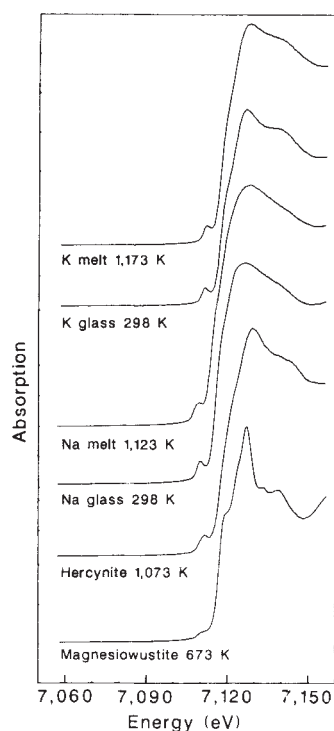


Fig. 4 Near-K-edge spectra (XANES) of the samples in Fig. 1. The $1s \rightarrow 3d$ feature is at $\sim 7,111$ eV. This feature has a maximum intensity of $\sim 6\%$ of the edge step in structures with mostly Fe^{2+} on a tetrahedral site, but 2% or less in all octahedral structures examined. Resolution is ~ 2.0 eV. Data recorded with silicon (111) monochromator crystals.

Mössbauer IS values, because such a highly distorted site would be much larger geometrically than, for example, the octahedral spinel Fe^{2+} sites and thus yield a larger IS. Hence both Mössbauer parameters and the XANES pre-edge feature intensity are consistent with the EXAFS results, and collectively verify Fe^{2+} in dominantly tetrahedral sites.

The Fe^{2+} in the silicate melt compositions in this study must therefore act as network formers, or belong to tetrahedral free-ion complexes. 'Modifier' sites are also a possibility, though with a smaller oxygen coordination than the 6–8-fold value that is usually assumed. Of the other two possibilities, the free-ion complex is more likely to have octahedral coordination, as Waseda¹ observed in melts with higher FeO and lower SiO_2 content. Furthermore, the stoichiometry of our melt samples is consistent with a complete network, with a ratio of tetrahedral network formers to oxygens near 1:2. Because of this the formation of free-ion complexes would necessitate large-scale charge imbalances throughout the melt, which are energetically unfavourable.

We therefore conclude that Fe^{2+} in our glasses and melts is a network former, contrary to popular assumption^{9,10}. This leads to the interesting possibility of pressure-induced changes from four- to six-coordination for Fe^{2+} in silicate melts in the upper mantle. This would be analogous to Waff's¹¹ proposed coordination change for Fe^{3+} , but more significant, as Fe^{3+} is a minor species by comparison. Such a coordination change would undoubtedly affect structure–property relationships for the melt, probably with changes in viscosity, density, compressibility and solid phases stable on the liquidus. Mg^{2+} is a more common ion in the upper mantle which, being somewhat smaller, is more likely than Fe^{2+} to enter tetrahedral coordination in melts. Mg^{2+} exists in tetrahedral coordination in some silicate crystals (such as melilite), and can be interpreted as four-coordinated in

MgSiO_3 glasses^{12–14}. Melt simulations^{15,16} have also predicted tetrahedrally coordinated Mg^{2+} . Our results on iron reinforce the predictions of tetrahedrally coordinated Mg^{2+} in melts and point to the possibility of pressure-induced coordination changes for Mg^{2+} in melts in the upper mantle.

This work was supported by NSF. The EXAFS/XANES analyses were carried out at the Stanford Synchrotron Radiation Laboratory, which is supported by DOE and NIH. We thank J. Stebbins for advice on glass preparation and M. Brandriss for assisting with the preparation and microprobe analyses.

Received 30 July 1987; accepted 7 January 1988.

- Waseda, Y. & Toguri, J. M. *Metall. Trans.* **9B**, 595–601 (1978).
- Lytle, F. W. *et al. Nucl. Instrum. Meth.* **226**, 542–548 (1984).
- Waychunas, G. A., Brown, G. E. Jr & Apter, M. J. *Phys. Chem. Miner.* **13**, 31–47 (1986).
- Hazen, R. M. & Finger, L. W. *Comparative Crystal Chemistry* (Wiley, New York, 1986).
- Eisenberger, P. & Brown, G. S. *Solid St. Commun.* **29**, 481–484 (1979).
- Simanek, E. & Sroubek, Z. *Phys. Rev.* **163**, 275–279 (1967).
- Dollase, W. A. *Phys. Chem. Miner.* **6**, 295–304 (1980).
- Waychunas, G. A., Apter, M. J. & Brown, G. E. Jr *Phys. Chem. Miner.* **10**, 1–9 (1983).
- Mysen, B. O., Virgo, D., Neumann, E.-R. & Seifert, F. A. *Am. Miner.* **70**, 317–331 (1985).
- Calas, G. & Petiau, J. in *The Structure of Non-Crystalline Materials* (eds Gaskell, P. H., Parker, J. M. & Davis, E. A.) 18–28 (Taylor and Francis, London, 1982).
- Waff, H. S. *Geophys. Res. Lett.* **2**, 193–196 (1975).
- Waseda, Y. & Toguri, J. M. *Metall. Trans.* **8B**, 563–568 (1977).
- Yin, C. D., Okuno, M., Morikawa, H. & Marumo, F. *J. non-cryst. Solids* **55**, 131–141 (1983).
- Sharma, S. & Yoder, H. S. *Yb. Carnegie Instn Wash.* **78**, 526–532 (1979).
- Matsui, Y. & Kawamura, K. *Nature* **285**, 648–649 (1980).
- Dempsey, M., Kawamura, K. & Henderson, C. *Prog. exp. Petrology* **6**, 57–59 (1985).
- Teo, B. K. & Lee, P. A. *J. Am. Chem. Soc.* **101**, 2815–2831 (1974).

Do lamprophyres carry gold as well as diamonds?

Nicholas M. S. Rock & David I. Groves

Department of Geology, University of Western Australia, Nedlands 6009, Australia

Diamonds are now known from kimberlites, lamproites and alkaline and ultramafic lamprophyres. Here we point out that these rocks are also significantly enriched in gold, as are calc-alkaline lamprophyres. The average gold abundance in lamprophyres seems to be at least an order of magnitude higher than in 'common' igneous rocks, and many individual values are 100–1,000 times higher. This high gold content may reflect two factors: (1) lamprophyres tap exceptionally deep regions of the Earth—regions not only where diamond is stable, but also where gold may be more enriched than in the source regions for other igneous rocks; (2) lamprophyre magmas are suitable carriers of gold from depth because, in their high CO_2 , H_2O , F, K, Rb and Ba but moderate S contents, they closely mirror fluids actually known to deposit gold veins in the crust. Most types of lamprophyre rise rapidly enough from great depths to retain any diamond and gold, along with rich mantle-xenolith suites. Calc-alkaline lamprophyres, on the other hand, undergo extensive crustal interactions which cause them to pick up crustal xenoliths, and to lose any diamond and mantle xenoliths along with at least some of their gold; this last may account for their association with mesothermal (for example, Archaean) gold deposits.

Lamprophyres are a diverse 'clan' of mafic to ultramafic, porphyritic, volatile-rich alkaline igneous rocks, here taken^{1,2} to include kimberlites³, as well as the more traditional⁴ groups of calc-alkaline lamprophyres (such as minettes)⁵, alkaline lamprophyres (such as monchiquites)², ultramafic lamprophyres (such as alnöites)⁶ and lamproites⁷. Their economic interest springs first from their association with diamonds (now confirmed in both monchiquites⁸ and ultramafic lamprophyres (refs 1, 6 and unpublished Australian company data), as well as in kimberlites and lamproites^{3,7,8}, and also from a spatial and temporal association between calc-alkaline lamprophyres and

Table 1 Some well-documented associations between gold and calc-alkaline lamprophyres

Area	Approx. age	References
Yilgarn, W. Australia	~2.7 Gyr	31, 32
Superior Province, Canada	~2.67 Gyr	13, 33
British Caledonides	400 Myr	32
Woods Point, Australia	380 Myr	34
Shandong, China	200 Myr	35
N. American Cordillera	Jurassic-Cretaceous/Tertiary	12, 36
Italian Alps	32 Myr	37

For full details, including evidence for exact contemporaneity and assessment of the structural/genetic significance of these associations, see ref. 9.

certain (especially mesothermal) types of gold deposit (see Table 1). The nature and significance of these gold-lamprophyre associations are discussed elsewhere⁹; our purpose here is to establish that lamprophyres seem to have intrinsically high Au contents, to present a hypothesis as to why this may be so, and to plead for more data to test this hypothesis.

Table 2 and Fig. 1 summarize the available data on Au in lamprophyres, incorporating 38 new analyses, and two other previously unpublished data sets. High Au enrichments have now been reported in all types of lamprophyre; the Au content of 'average' lamprophyre (represented by mean, median or mode) is apparently one or two orders of magnitude higher than in common igneous rocks (~1–2 p.p.b. Au)^{10,11}, and the long tail to the right in Fig. 1 indicates the many individual lamprophyres that are even more highly enriched. Since in some gold provinces, potential source rocks with only 5–10 p.p.b. Au are economically interesting¹², these data would suggest that lamprophyres could be exploration targets in their own right.

As many putative 'gold-rich rocks' have proved contentious or spurious^{10,11}, two problems must be immediately addressed regarding these data. Firstly, older analytical methods tend to give higher Au values at low concentrations than the most modern methods (based on instrumental neutron activation analysis). We doubt that this is a problem here, as only 6 of the 84 individual values shown in Fig. 1 were obtained by older techniques. Secondly, the high-Au calc-alkaline lamprophyres in Table 2 are in fact associated with Au mineralization, and it could be argued¹⁰ that they have all been secondarily mineralized. We doubt that this is so for three reasons. (1) Nearly all the samples analysed in Table 2 were specifically chosen as petrographically fresh, and show no visible signs of characteristic gold-related alteration. (2) "Altered" and "least-altered" lamprophyres, where separated, have statistically indistinguishable Au contents¹³ (Table 2); there is thus no evidence that alteration has enhanced Au beyond 'primary' values, although this conclusion must be tentative with so few data. (3) High Au abundances also occur in kimberlites, lamproites and alkaline and ultramafic lamprophyres (Table 2), which are nowhere known to be coeval with gold deposits; indeed, there is no group of igneous rocks less likely to suffer crustal-style gold alteration than kimberlites and lamproites, which can ascend extremely rapidly into the crust from exceptional depths, bringing diamonds and mantle xenoliths with them³.

That high Au contents may be an intrinsic feature of any lamprophyre is understandable in view of the following characteristics of lamprophyres. First, they may be eminently suitable carriers of gold. Most workers agree¹⁴ that, to form 'gold-only' crustal deposits, Au is transported in low salinity, CO₂-rich, moderate-S fluids, which complex Au as, for example, HAu(HS)₂, but cannot transport base metals¹⁵. Lamprophyres seem to approach these fluids extremely closely: they have moderate S contents, coupled with the highest contents among igneous rocks of H₂O, CO₂, F, K, Ba, Rb and related elements^{1–8}—precisely those elements enriched in alteration halos around gold deposits. The local association of F-rich fenites

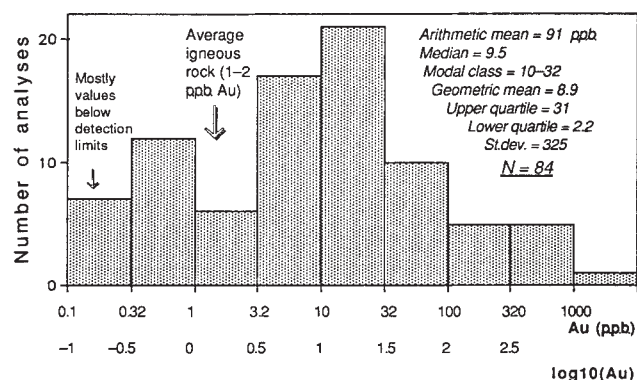


Fig. 1 Histogram of individual Au analyses for all types of lamprophyre (sources listed in Table 2). Note the logarithmic x-axis. Of the 84 values incorporated, 78 were obtained by the most modern analytical techniques, and should be within 10–20% of the true values. Values below detection limits are set to an arbitrary value of 0.1 p.p.b. Averages are incorporated as single data values. The distribution is log-normal ($\sqrt{b_1}$ (skew) and b_2 (kurtosis) of log-values are -0.002 and -0.1 respectively), as already suggested by the near identity of the median and geometric mean.

with both lamprophyres^{1,6} and gold deposits¹⁶ further indicates a close similarity between lamprophyric fluids and those depositing hydrothermal gold veins. Next, lamprophyres are low-volume extracts under volatile-rich conditions from fertile, modally metasomatized mantle at extreme depths (>150 km). Although we know of no Au analyses for xenoliths representing the type of metasomatized mantle envisaged (mantle with added phases such as phlogopite, hollandite minerals and possible carbonates¹⁷), Au-rich sulphide fractions and intergranular material are already known from 'ordinary' lherzolite xenoliths and, significantly, are assumed to be derived from metasomatizing fluids^{18,19}. Low-volume lamprophyre melts rich in H₂O, CO₂, F, K, Ba and Rb are suitable vehicles for extracting any of this intergranular Au. Finally, the ultimate sources of lamprophyres may be Au-enriched. The metasomatized mantle source for lamprophyres may itself be generated by thermal or chemical plumes rich in K, Ba, Sr and Zr, rising from as deep as the core-mantle boundary^{20,21}. Although we would not wish to lay great stress on aspects as speculative as involvement of the core, this has long been assumed to be the most Au-rich part of the Earth, reflecting gold's density and its combined siderophile and chalcophile characteristics^{10,19,22}. Typical estimates of the Au content of the core reach ~500 p.p.b.—several hundred times average mantle or crustal values^{10,22}—and are supported (in a chondritic Earth model) by very high Au contents in chondrites (average, 222 p.p.b.; maximum 8,700 p.p.b.)¹⁰. If such plumes do operate, they may provide a unique vehicle for extracting Au metasomatically from the core, enhancing it in the mantle zones that they metasomatize, and hence generating Au-rich lamprophyres by melting of those zones. In summary, conditions appear ideal, at every stage in the genesis of lamprophyres, for Au to be extracted from deep sources and transported efficiently upwards into the crust.

Assuming, then, that all lamprophyres may potentially be Au-bearing, why are only calc-alkaline lamprophyres intimately associated with gold deposits? This can be answered by contrasting them with other lamprophyres^{2,3}. Whereas other lamprophyres contain diamonds (indicating source depths of >150 km), are rich in mantle xenoliths and display an unequivocally mantle-type geochemistry, the calc-alkaline lamprophyres contain no diamonds, are rich in crustal xenoliths, and their geochemistry shows 'crustal' fingerprints⁵. Very rare mantle xenoliths indicate source depths of >130 km, however²³. Fur-

Table 2 Au data for all types of lamprophyre

Area	Lamprophyre type(s)*	No. of analyses	Au content (p.p.b.)			Analysis method	Reference
			Mean	Median	Max.		
N. Muratau Mtns, USSR	Not specified	1	—	—	20	?	38†
S. Donetz Basin, USSR	Alkaline (camptonite, monchiquite)	6	25	20	80	?	39†
India & southern Africa	Kimberlite	11	12	10	103	INAA	30‡
Yilgarn, W. Australia	Ultramafic (alnöite, aillikite, etc.)	5	45	12	180	Fire assay	Unpubl. data
Kimberley, W. Australia	Lamproite	5	25	20	54	ICPMS	40‡
Canadian Shield	? Mainly calc-alkaline	15	1.6	?	2.0	Fire assay/AAS	10 (p. 38)†§
Canadian Arrow, Canada	Calc-alkaline	8	683	380	2,700	?	13§
Shandong, China	Calc-alkaline	?	?	?	25	?	35§
SW Scotland	Calc-alkaline	9	99	137	523	Fire assay/INAA	32‡
Yilgarn, W. Australia	Calc-alkaline	38	7	3	112	Fire assay/INAA	New data‡
Global	All lamprophyres	84	91	9.5	2,700		

INAA, instrumental neutron activation analysis; ICPMS, inductively coupled plasma mass spectrometry.

* 'Lamprophyre' refers to a clan of rock families, covering calc-alkaline, alkaline and ultramafic lamprophyres, kimberlites and lamproites^{1,2}.

† Analyses by traditional techniques which may not be fully compatible with more modern data.

‡ Should be within 10–20% of true values.

§ Three analyses of 'least-altered' lamprophyres (range 14–620 p.p.b. Au) and five of 'altered lamprophyres' (69–2,700 p.p.b.); nonparametric tests (Mann-Whitney, Kolmogorov-Smirnov, Wald-Wolfowitz runs) fail to reject the null hypothesis that these least-altered and altered lamprophyres come from the same population as regards Au content.

|| New data mostly by X-ray laboratories, Don Mills, Ontario, Canada, run together with 'blind' blanks (pure quartz) and standards (Canadian gold ore MA2) to ensure analytical precision and accuracy; samples were all chosen as the most petrographically fresh available, and none are known to have suffered mineralization.

thermore, calc-alkaline lamprophyres show a spatial, temporal and genetic association with granitoids and porphyries⁵; such an association is unknown for the other types for lamprophyre.

Evidence is now accumulating that calc-alkaline lamprophyres may represent crustally contaminated lamproite magmas^{5,24}, and that they may give rise to andesitic to rhyolitic porphyries and certain granitoids by differentiation and crustal interaction processes^{25–29}. Differences in Au content may therefore reflect differences in the extent of crustal interaction experienced by different lamprophyres during their ascent from depth. Kimberlites and lamproites ascend sufficiently rapidly to retain not only their mantle xenoliths and diamonds, but also their mantle-type chemistry and inherited Au (there is no opportunity for it to be released). On the other hand, by interacting strongly with the crust, calc-alkaline lamprophyres lose any diamond and mantle xenolith suites they bring from greater depths, gain abundant crustal xenolith suites, and may release part or all of their inherited Au into crustal outgassing or metamorphic systems⁹.

The above ideas by no means require every lamprophyre to be Au-rich. Mantle metasomatism is notoriously heterogeneous¹⁷, and Au-poor lamprophyres may merely mirror barren mantle sources; alternatively, calc-alkaline lamprophyres may have given up their Au to form deposits. Finally, barren lamprophyres may dominate Au-bearing examples just as barren kimberlites and lamproites dominate diamond-bearing ones^{3,7}. The single set of consistently low-Au calc-alkaline lamprophyres in Table 2 (Canadian Shield) and the rather more erratic distribution of values for the calc-alkaline lamprophyres as a whole, support these ideas.

If substantiated by further data, high-Au lamprophyres would have major implications not merely for theoretical modelling of crustal and mantle processes, but perhaps also for understanding the nature of the mantle and core themselves. The apparent combination of high platinum-group-element (PGE) with high Au in kimberlites³⁰ raises even more intriguing possibilities (although no PGE data for other lamprophyres are available). Meanwhile, these ideas may affect exploration, where the difference between a kimberlite, a madupite, a minette and a monchiquite may yet prove to be that between a new diamond mine, a future diamond mine, a gold mine, or no mine at all.

Received 17 June 1987; accepted 20 January 1988.

1. Rock, N. M. S. in *Proc. 4th int. Kimberlite Conf.* (Geological Society of Australia, in the press).
2. Rock, N. M. S. *Spec. Publs. geol. Soc. Lond.* **30**, 191–226 (1987).
3. Mitchell, R. H. *Kimberlites: Mineralogy, Geochemistry and Petrology* (Plenum, London, 1986).
4. Streckeisen, A. *Geology* **7**, 331–335 (1979).
5. Rock, N. M. S. *Trans. R. Soc. Edinb.* **74**, 193–227 (1984).
6. Rock, N. M. S. *J. Petrology* **27**, 155–196 (1986).
7. Bergman, S. C. *Spec. Publs. geol. Soc. Lond.* **30**, 103–190 (1987).
8. Jaques, A. L., Kerr, I. D., Lucas, H., Sun, S.-S. & Chappell, B. W. in *Proc. 4th int. Kimberlite Conf.* (Geological Society of Australia, in the press).
9. Rock, N. M. S. & Groves, D. I. *Geology* (in the press).
10. Boyle, R. W. *Bull. geol. Surv. Can.* **280** (1979).
11. Meyer, M. & Saager, R. *Miner. Deposita* **20**, 284–289 (1985).
12. Mutschler, F. E., Griffen, M. E., Stevens, D. S. & Shannon, S. S. *Trans. geol. Soc. S. Afr.* **88**, 355–378 (1985).
13. McNeil, A. M. & Kerrich, R. *Can. J. Earth Sci.* **23**, 324–343 (1986).
14. Kerrich, R. in *Proc. Gold Workshop, Yellowknife* (ed. Morton, R. D.) 95–171 (University of Alberta, Edmonton, 1979).
15. Phillips, G. N. & Groves, D. I. *J. geol. Soc. Aust.* **30**, 25–39 (1983).
16. King, R. W. & Kerrich, R. *Can. J. Earth Sci.* **24**, 479–502 (1987).
17. Menzies, M. A. & Hawkesworth, C. J. *Mantle Metasomatism* (Academic, London, 1987).
18. Frey, F. A. & Green, D. H. *Geochim. cosmochim. Acta* **38**, 1023–1059 (1974).
19. Mitchell, R. H. & Keays, R. R. *Geochim. cosmochim. Acta* **45**, 2425–42 (1981).
20. Edgar, A. D. *Neues Jb. Miner. Abh.* **147**, 35–46 (1983).
21. Morgan, W. J. *Mem. geol. Soc. Am.* **132**, 7–22 (1972).
22. Wedepohl, K. H. *Handbook of Geochemistry* (Springer, Berlin, 1978).
23. Ehrenberg, S. N. in *The Mantle Sample* (eds Boyd, F. R. & Meyer, H. A. O.) 330–344 (American Geophysical Union, Washington, 1979).
24. Foley, S. F., Green, D. H., Venturelli, G. & Toscani, L. *Earth Sci. Rev.* **24**, 81–134 (1987).
25. Rock, N. M. S. & Hunter, R. H. *Geol. Rdsch.* **76**, 805–826 (1987).
26. Leat, P. T., Thompson, R. N., Morrison, M. A., Hendry, G. L. & Travhorn, S. C. *Trans. R. Soc. Edinb.* (in the press).
27. Macdonald, R., Rock, N. M. S., Rundle, C. C. & Russell, O. J. *Mineralog. Mag.* **50**, 547–557 (1986).
28. Suzuki, K. & Shiraki, K. *Contr. Miner. Petrol.* **71**, 313–322 (1980).
29. Thompson, R. N. & Fowler, M. B. *Contr. Miner. Petrol.* **74**, 507–522 (1986).
30. Paul, D. K., Crockett, J. H. & Nixon, P. H. in *Kimberlites, Diatremes and Diamonds* (eds Boyd, F. R. & Meyer, H. A. O.) 272–279 (American Geophysical Union, Washington, 1979).
31. Hallberg, J. A. *Geology and Mineral Deposits of the Leonora-Laverton Area* (Hesperian, Perth, 1985).
32. Rock, N. M. S., Duller, P., Haszeldine, R. S. & Groves, D. I. *Geol. Dep. Univ. Extension, Univ. West Australia, Publ.* **11**, 271–286 (1987).
33. Wyman, D. & Kerrich, R. *Econ. Geol.* (in the press).
34. Hills, E. S. in *Sir Douglas Mawson Anniversary Volume* (eds Glaessner, M. F. & Rudd, E. A.) 87–100 (University of Adelaide, 1952).
35. Liu, L., Kong, Q., Yao, Y. & Gong, R. *J. Changchun Coll. Geol.* **38**, 13–27 (1984).
36. Drysdale, C. W. *Mem. geol. Surv. Can.* **77** (1915).
37. Diamond, L. W. & Weidenbeck, M. *Schweiz. miner. petrogr. Mitt.* **66**, 385–394 (1986).
38. Buturlinov, N. V. & Latysh, I. K. *Geologicheskij Zhurnal* (Kiev) **30**, 69–75 (1970).
39. Kushmuradov, O. K. *Uzbek. geol. Zh.* **14**, 8–11 (1970).
40. Lewis, J. D. M.Sc. thesis, Univ. W. Australia (1987).

A comparative embryological study of two ornithischian dinosaurs

John R. Horner & David B. Weishampel*

Museum of the Rockies, Montana State University, Bozeman, Montana 59717, USA

* Department of Cell Biology and Anatomy, The Johns Hopkins University, School of Medicine, Baltimore, Maryland 21205, USA

Embryonic dinosaur remains are extremely rare^{1,2} and with the exception of the specimens described here, are so fragmentary that they have yielded little or no information useful for palaeobiological interpretation. We report here the first well-documented association of dinosaur eggs containing embryos, juveniles and adults from the fossil record. The material, attributable to a new hypsilophodontid, *Orodromeus makelai*, n. gen., n. sp., and the hadrosaurid *Maiasaura peeblesorum*, comes from the Upper Cretaceous Two Medicine Formation of western Montana³. Osteological and histological studies of these two animals indicate structural and morphological adaptations for their respective growth patterns and neonate behaviour.

The hypsilophodontid embryonic skeletons (Fig. 1a) were found in a clutch of 19 unhatched eggs. The embryonic hadrosaurid skeletal elements were found as disarticulated elements strewn through nest sediments containing abundant, large eggshell fragments known from other nests containing young⁴ to belong to *M. peeblesorum*.

All of the skeletal elements present in the adult hypsilophodontids were ossified in the embryos, although some processes such as the fourth trochanters of the femura are absent and may not have ossified until after the muscles attached to these processes became active. The teeth which had erupted (Fig. 1b) are enamelled but lack the denticles found on the anterior and posterior borders of the crowns of older individuals (Fig. 2b). The vertebrae are ossified, but very spongy. Limb-element ossification varies in both proximal and distal epiphyseo-metaphyseal regions, and shows remarkable differences between *O. makelai* and *M. peeblesorum*, which are most prob-

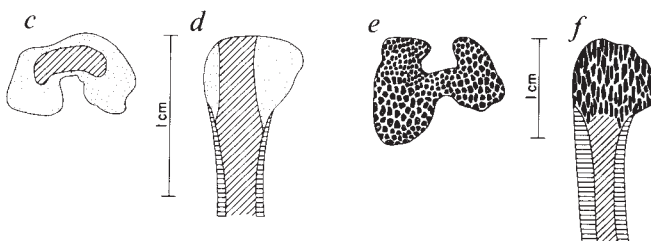
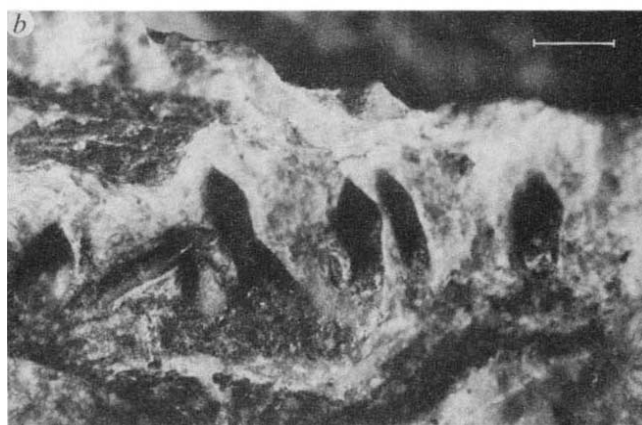
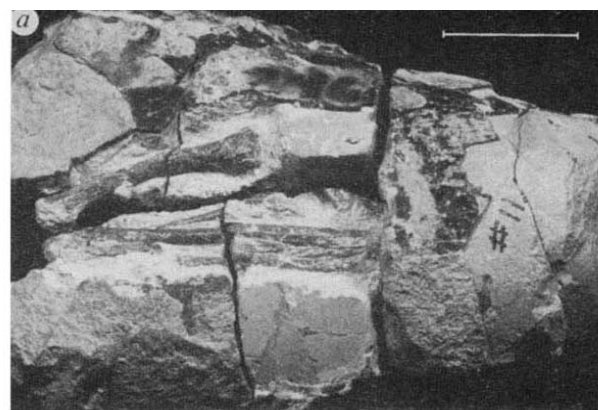


Table 1 Comparative measurements of the hindlimb in hypsilophodontid dinosaurs

	Femur length (mm)	Tibia length (mm)	MT III length (mm)
<i>Orodromeus makelai</i> (MOR 294)	106	132	—
<i>Orodromeus makelai</i> (PU 22412)	118	155	65.5
<i>Orodromeus makelai</i> (MOR 246-11)	36	48	27
<i>Dryosaurus altus</i> (YPM 1876)	360	400	—
<i>Dryosaurus lettowvorbecki</i> (HMV dy VI)	135	152.5	76
<i>Hypsilophodon foxii</i> (BMNH R5830)	101	117	62.5
<i>Nanosaurus rex</i> (BYU ESM-163R)	151	180	83
<i>Parksosaurus warreni</i> (ROM 804)	270	300	151
<i>Thescelosaurus neglectus</i> (USNM 7757)	335	300	106
<i>Yandusaurus multidentis</i> ⁹	151	187	83

Abbreviations: BMNH, British Museum (Natural History), London, England; BYU, Brigham Young University Earth Science Museum, Provo, Utah, USA; HMN, Museum für Naturkunde, Humboldt Universität, East Berlin, GDR; MOR, Museum of the Rockies, Bozeman, Montana, USA; NMC, National Museums of Canada, Ottawa, Ontario, Canada; PU, Princeton University (collections now housed at Yale University, Peabody Museum, New Haven, Connecticut, USA); ROM, Royal Ontario Museum, Toronto, Ontario, Canada; USNM, United States National Museum of Natural History, Washington, DC, USA; YPM, Yale Peabody Museum, New Haven, Connecticut, USA.

Fig. 1 a, MOR 246-11, view of articulated left leg of embryonic *Orodromeus makelai* in situ, within egg. b, Anterior portion of right maxilla showing teeth (embryonic *O. makelai* MOR 246-11). c, View of epiphyseal surface of distal femur (embryonic *O. makelai* MOR 246-1). d, Sagittal section of distal femur (embryonic *O. makelai* MOR 246-1) in lateral view showing canal opening. e, View of epiphyseal surface of distal femur (embryonic *Maiasaura peeblesorum*, PU 22432). f, Sagittal section of distal femur (embryonic *M. peeblesorum*, PU 22432) showing endochondral metaphysis. Scale bar: a, 2 cm; b, 1 mm.

ably linked to growth rates and behavioural constraints. The proximal and distal ends of the embryonic femur of *O. makelai* (Fig. 1c, d) possess well formed, smooth condyles which, although fully ossified in appearance, are formed entirely of calcified cartilage. Endochondral bone is not observed in the epiphyseal or metaphyseal regions. Canals penetrate the centre of the epiphyseal surfaces and extend into the medullary cavity. These openings would have bounded the cartilage cones that extended from the medullary cavity to the epiphyseal surface. The presence of well-formed calcified cartilage condyles may indicate that the cartilaginous epiphysis was relatively small.

In contrast, the epiphyseal surfaces of the *M. peeblesorum* femur (Fig. 1e, f) are composed of a thin layer of calcified

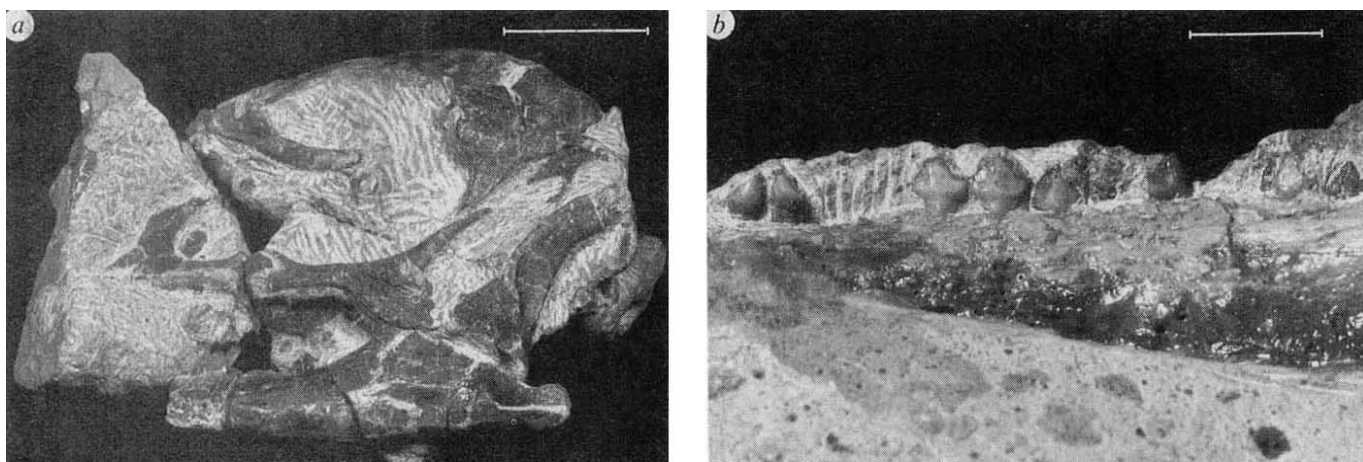


Fig. 2 MOR 294, *Orodromeus makelai* Holotype skull, left lateral view. b, *O. makelai* MOR 248, left lateral view of left dentary showing teeth. Scale bar: a, 2 cm; b, 5 mm.

cartilage overlying a very spongy endochondral metaphysis. Only the most proximal portions of the condyles exist and the ends of the bones therefore appear spongy and incomplete. Well formed condyles, composed of a thin outer layer of calcified cartilage and inner endochondral bone, are found on limb bones of individuals with lengths of ~1 m. No open canals exist in the epiphyseal ends of the embryos or young, although a massive cartilage cone presumably existed and extended into the medullary cavity via the spaces within the endochondral lattice, as occurs in living reptiles and birds. The embryonic and juvenile compact periosteal bone of limb elements of both species is highly vascularized, fibrous and plexiform in structure.

Previous discoveries have suggested that *M. peeblesorum* was nest-bound during the period of time between hatching and attaining a length of ~1.5 m (refs 3, 4); adults grew to an approximate length of 8 m. *O. makelai* was apparently precocial upon hatching, and attained an adult length of ~2.5 m.

As is the case with altricial birds, growth was probably the chief demand on hadrosaurid nestling metabolism and the nestlings invested little energy to shape condyles until they were nearing the time to leave the nest. The hypsilophodontid, by contrast, did not have prolonged nesting behaviour, nor did adults attain large size; juveniles would have required well-formed bones and epiphyseal condyles to have enabled them to perform similar activities as their adult counterparts.

Genus *Orodromeus* nov.

Derivation of name: *Oros* (Greek): mountain; *Dromeus* (Greek): runner. The name alludes to 'Egg Mountain,' as well as the state of Montana, and to the animals, presumed cursorial habits.

Type species: *O. makelai*.

Diagnosis: As for species *O. makelai*, see below.

Species *Orodromeus makelai* nov.

Derivation of name: Species, *makelai* to honour the late Robert Makela for his many dinosaur discoveries including the holotype.

Holotype: MOR294, virtually complete skull (Fig. 2a) and skeleton (without tail).

Referred material: MOR 246, clutch of 19 eggs with embryos; PU 22412, hind legs; MOR 331, skeleton; MOR 248, skull and skeleton; MOR 403, braincase.

Locality: Egg Mountain site, Willow Creek Anticline, Teton County, Montana, USA.

Horizon: Upper Two Medicine Formation, Upper Cretaceous (? Late Campanian).

Diagnosis: Palpebral anchored to the postorbital part of the posterior margin of orbit. Lateral boss on the jugal. Maxillary and dentary teeth broadly triangular in lateral view, often as wide as they are high. Rounded apical (primary) ridge subequally dividing anterior and posterior region of the crown. Anterior and posterior edges of the crown bear five to six uniformly-sized denticles in adult. Radiale, intermedium, and ulnare, plus one distal carpal form unfused wrist.

Orodromeus makelai represents the first known hypsilophodontid from the Campanian of North America. Hindlimb proportions, from embryonic stages through adulthood, indicate that *O. makelai* was an extremely gracile and fast running animal, much more so than other hypsilophodontids with the exception of *Yandusaurus multidentis* from the People's Republic of China (Table 1). *O. makelai* is typically hypsilophodontid in all characters except its dentition (Fig. 2b). Here it resembles very closely the condition in primitive ornithischians⁵. The tooth crowns are extremely reminiscent of *Lesothosaurus diagnosticus*. Although the configuration of the skull in *O. makelai* suggests the presence of pleurokinesis⁶⁻⁸, to produce transversely oriented mastication, tooth wear indicates reversion to (or retention of) the primitive high-angle, fabrosaur-like style of chewing.

We thank Karen Chin for photographs, Kris Ellingsen for drawings and David Norman and Andrew Milner for their helpful reviews. This work was supported by the NSF.

Received 11 September 1987; accepted 25 January 1988.

1. Sochava, A. V. *Paleont. Zhur.* (4), 82-92 (1972).
2. Kitching, J. W. *Paleont. Afr.* 22, 41-45 (1979).
3. Horner, J. R. *Scient. Am.* 250, 130-137 (1984).
4. Horner, J. R. & Makela, R. *Nature* 282, 296-298 (1979).

5. Galton, P. M. *Palaeont. Z.* 52, 138-159 (1978).
6. Norman, D. B. *Symp. Zool. Soc. Lond.* 52, 521-547 (1984).
7. Weishampel, D. B. *Adv. Anat. Embryol. Cell Biol.* 87, 1-110 (1984).
8. Norman, D. B. & Weishampel, D. B. *Am. Nat.* 126, 151-164 (1985).
9. He, X.-L. & Cai, K.-J. *J. Chengdu College Geol. Suppl.* 1, 5-14 (1983).

Phylogenetic trees support the coevolution of parasites and their hosts

Mark S. Hafner & Steven A. Nadler

Museum of Zoology and Department of Zoology and Physiology,
Louisiana State University, Baton Rouge, Louisiana 70803, USA

The close correspondence often observed between the taxonomy of parasites and their hosts^{1,2} has led to *Fahrenholz's rule*³, which postulates that parasites and their hosts speciate in synchrony. This leads to the prediction that phylogenetic trees of parasites and their hosts should be topologically identical^{2,4}. We report here a test of this prediction which involves the construction of phylogenetic trees for rodents and their ectoparasites using protein electrophoretic data. We find a high degree of concordance in the branching patterns of the trees which suggests that there is a history of cospeciation in this host-parasite assemblage. In several cases where the branching patterns were identical in the host and parasite phylogenies, the branch lengths were also very similar which, given the assumptions of molecular clock theory, strongly suggests that the speciation of these hosts and ectoparasites was roughly contemporaneous and causally related.

We have performed a direct test of *Fahrenholz's rule* by comparing independent phylogenies based on biochemical data for a group of rodents and their ectoparasites. The rodents in this study include eight species of pocket gophers, representing three genera in the rodent family Geomyidae; these fossorial herbivores are usually solitary, and geomyid species are generally allopatric. The ectoparasites include ten species of chewing lice representing two genera in the mallophagan family Trichodectidae. The life cycle of these wingless insects occurs entirely on the host and includes three principal stages: egg, nymph, and adult. Chewing lice have a generation time of ~40 days⁵, which is roughly one-fifth the generation time of pocket gophers⁶. Transmission of chewing lice among pocket gophers is thought to occur only through host-to-host contact⁷, and the combination of low parasite vagility and obligate contact-transmission of lice should limit opportunities for colonization of new host species. The absence of widespread transfer of lice among host species should, in turn, increase the likelihood of detecting cospeciation in this host-parasite assemblage.

Using standard electrophoretic procedures⁸, we surveyed 31 protein loci in pocket gophers and 14 loci in chewing lice brushed from the pelage of the pocket gophers. To ensure that our electrophoretic survey of lice was not an examination of pocket gopher proteins contained in the gut of the louse, we compared protein variation at the ten loci in common to the two surveys. At all ten loci, hosts and parasites showed different electrophoretic patterns and different patterns of variation (Fig. 1). Each data set was clustered using both cladistic (locus-by-locus^{9,10}) and phenetic¹¹ procedures; tree topology was unaffected by the clustering method used.

The host and parasite trees are topologically identical in all but three regions (indicated by daggers in Fig. 2). In most cases, sister taxa of lice parasitize hosts that are also sister taxa (see nodes A and F, Fig. 2), and branching sequences above the species level in the two groups are identical (nodes B-E). The probability of this level of topological similarity occurring by chance alone was calculated using the component-replication method of Nelson and Platnick¹². This method requires equal numbers of taxa in the trees being compared, making it necessary for us to reduce the number of taxa in our louse tree. Because each species of *Thomomys* hosts two species of lice (Fig. 2), we first calculated probability levels excluding the two *Thomomys* species, *wardi* and *minor*. The probability of chance similarity of the two trees (without *wardi* and *minor*)

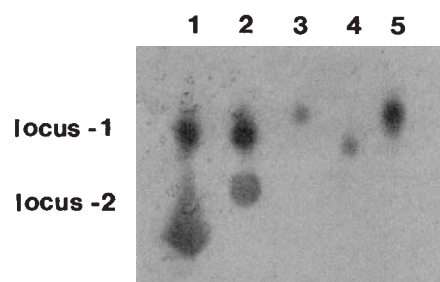


Fig. 1 An example of electrophoretic variation at one of ten protein loci (isocitrate dehydrogenase-1) examined in both pocket gophers and lice. Pocket gopher samples are in lanes 1 and 2, lice in lanes 3-5. Allele designations at locus 1 are listed after species names: lane 1, *Orthogeomys heterodus* A; lane 2, *Thomomys bottae* A; lane 3, *Thomomys minor* b; lane 4, *Geomys actuosus* d; lane 5, *Geomys costaricensis* c. The pocket gopher in lane 1 hosts the louse species in lane 5, and the lice in lanes 3 and 4 coexist on the host in lane 2 (see Fig. 2). Only one isocitrate dehydrogenase locus was evident in lice.

was low ($P=0.05$). However, when *wardi* and *minor* were included and the two *Geomys* species (*thomomys* and *actuosus*) excluded, the probability of chance similarity of the host and parasite trees was remote ($P=0.001$). This high degree of tree matching is consistent with predictions of the cospeciation hypothesis.

The three host-parasite associations indicated by asterisks in Fig. 2 probably result from host switching (lateral transfer) by lice. In these cases, the phylogenetic history of the louse taxon does not mirror that of its host, thereby falsifying the cospeciation hypothesis for these (and only these) associations. It is important to note that no case of suspected host switching involves hosts that are geographically disjunct; in each case, the geographical range of the colonizer's current host abuts that of the colonizer's putative ancestral host (that is, the pocket gopher parasitized by the colonizer's sister taxon)¹³.

Not only does the cospeciation hypothesis predict similar patterns of speciation in host-parasite assemblages, but it also predicts that the speciation events were causally related and, therefore, approximately contemporaneous. Thus, we can test the cospeciation hypothesis from another perspective if we assume that proteins evolve in a roughly clock-like fashion¹⁴. If hosts and parasites are actually speciating in synchrony then, according to the 'molecular clock' hypothesis^{15,16}, we should expect to see roughly equivalent amounts of protein change in associated host and parasite lineages following speciation events (represented by nodes on phylogenetic trees). Therefore, to be consistent with the strict cospeciation model, host and parasite phylogenies must agree not only in branching pattern, but also in branch lengths, which are proportional to amounts of protein change in Fig. 2. Estimates of genetic distance, hence branch lengths, may be influenced by the ratio of slow:fast evolving loci surveyed¹⁷; in this study, 4 of 31 loci (13%) examined in pocket gophers, and 2 of 14 loci (14%) surveyed in chewing lice, can be classified as fast-evolving¹⁸.

Analogous nodes on the two trees (A-F in Fig. 2) represent hypothesized cospeciation events in the history of the assemblage. Four of these nodes (A, C, D, E) are positioned at very similar genetic distances in the two trees ($\leq 4\%$ difference in each case), and we interpret this concordance as further corroboration of the cospeciation hypothesis. Specifically, we suggest that the speciation events represented by nodes A, C, D and E were approximately contemporaneous, and that the host and parasite lineages involved have accumulated protein differences at equal rates.

The dissimilarity between host and parasite genetic distances at nodes F in Fig. 2 suggests that the speciation events represented by these nodes were not contemporaneous (in that speciation

The figure displays three phylogenetic trees. The first tree, labeled 'Pocket gophers', shows relationships between *Thomomys* (talpoides, bottae), *Geomys* (bursarius), and *Orthogeomys* (hispidus, yucatanensis, cavator, underwoodi, cherriei, heterodus). The second tree, labeled 'Host-parasite associations', shows relationships between *Thomomys* (wardi, minor, thomomys, actuosi) and *Geomys* (ewingi). The third tree, labeled 'Chewing lice', shows relationships between *Thomomydoecus* and *Geomydoecus*. Both the host-parasite and chewing lice trees include asterisks (*) indicating specific associations. The trees are rooted at the bottom and branch upwards. A scale bar at the bottom indicates 'Rogers' distance' from 0.0 to 0.8.

The use of biochemical genetic evidence to investigate the evolutionary histories of hosts and their parasites allows us to view coevolution from two perspectives: sequence and timing of phylogenesis. Data are gathered independently for hosts and parasites, thus avoiding the nagging problem of circularity in studies of host-parasite coevolution. Patterns of allele-sharing reveal the sequence of phylogenetic events in each group, and comparisons of genetic distances between associated host and parasite lineages provides a single metric to assess the temporal relationship of speciation events in both. Using biochemical methods, we have conducted rigorous tests of the cospeciation

We thank D. J. Hafner, R. D. Price and R. M. Sullivan for assistance. D. W. Foltz, D. P. Pashley and J. V. Remsen Jr provided helpful criticisms of an earlier draft of this manuscript. This work was supported by a grant from the NSF.

1. Mitter, C. & Brooks, D. R. in *Coevolution* (eds Futuyma, D. J. & Slatkin, M.) 65-98 (Sinauer, Sunderland, Massachusetts, 1983).
2. Stone, A. R. & Hawksworth, D. L. *Coevolution and Systematics* (Clarendon, Oxford, 1986).
3. Eichler, W. *Annls mag. nat. Hist.* **12**, 588-598 (1948).
4. Lyal, C. H. C. *J. nat. Hist.* **21**, 1-28 (1987).
5. Marshall, A. G. *The Ecology of Ectoparasitic Insects* (Academic, London, 1981).
6. Smolen, M. J., Genoways, H. H. & Baker, R. J. *J. Mammal* **61**, 224-236 (1980).
7. Timm, R. M. in *Coevolution* (ed. Nitecki, M. H.) 225-265 (University of Chicago Press, 1983).
8. Harris, H. & Hopkinson, D. A. *Handbook of Enzyme Electrophoresis in Human Genetics* (North-Holland, Amsterdam, 1976).
9. Hafner, M. S., Hafner, J. C., Patton, J. L. & Smith, M. F. *Syst. Zool.* **36**, 18-34 (1987).
10. Honeycutt, R. L. & Williams, S. L. *J. Mammal* **63**, 208-217 (1982).
11. Sneath, P. H. A. & Sokal, R. R. *Numerical Taxonomy* (Freeman, San Francisco, 1973).
12. Nelson, G. & Platnick, N. *Systematics and Biogeography* (Columbia University Press, 1981).
13. Hall, E. R. *The Mammals of North America* (Wiley, New York, 1981).
14. Wilson, A. C., Carlson, S. S. & White, T. J. *A. Rev. Biochem.* **46**, 573-639 (1977).
15. King, J. L. & Jukes, T. H. *Science* **164**, 788-798 (1969).
16. Fitch, W. M. in *Molecular Evolution* (ed. Ayala, F. J.) 160-178 (Sinauer, Sunderland, Massachusetts, 1976).
17. Sarich, V. M. *Nature* **265**, 24-28 (1977).
18. Kojima, K., Gillespie, J. H. & Tobari, Y. *N. Biochem. Genet.* **4**, 627-637 (1970).
19. Rogers, J. S. *Univ. Texas Publ.* **72**, 145-153 (1972).

Control of nitrogen fixation by oxygen depletion in surface-associated microzones

Hans W. Paerl* & Richard G. Carlton†

* Institute of Marine Sciences, University of North Carolina, Chapel Hill, Morehead City, North Carolina 28557, USA

† W. K. Kellogg Biological Station, Michigan State University, Hickory Corners, Michigan 49060, USA

Chronic deficiencies in biologically utilizable nitrogen control primary production in vast segments of the world's oceans¹⁻³. *A priori*, it would appear that these conditions offer selective advantages to N₂-fixing microorganisms (eubacteria and cyanobacteria). Most studies, however, concur that ecologically insignificant rates of N₂ fixation are found in such waters⁴⁻⁶. We have addressed this paradox by identifying some environmental factors that may control the development and growth of marine N₂ fixers and experimentally manipulating them in nitrogen-depleted North Carolina coastal Atlantic Ocean water. The availability of phosphorus, iron and molybdenum, required for both synthesis and function of the N₂-fixing enzyme complex nitrogenase⁷, was sufficient to support the growth of N₂-fixers. However, colonizable inorganic and organic surfaces were important for the development and proliferation of N₂-fixing microbial consortia (mixtures of eubacteria, and cyanobacteria plus eubacteria). Surface-associated N₂ fixation potentials were much more dependent on organic than on inorganic nutrient enrichment, and were affected by the incidence and magnitude of localized O₂ depletion in the surface-associated microzones. As molecular O₂ is a potent inhibitor of N₂ fixation in many marine microorganisms, the extent to which biological nitrogen demands can be met by N₂ fixation depends on the presence and maintenance of O₂ depleted microzones.

Nutrient analyses and nutrient addition bioassays have characterized North Carolina's full salinity (32-34 ppt) Atlantic Coastal waters as chronically nitrogen-limited all year round⁸⁻¹⁰. Near-surface water samples were routinely collected during 1984-1987 at a location 1 km outside Beaufort Inlet, North Carolina, in open Atlantic Ocean waters, using deionized water rinsed pre-sterilized 20-l polypropylene carboys. Bioassay experiments were conducted on freshly sampled water to assess N₂ fixation potentials in untreated water and samples with nutrients and/or particles or surfaces added. Pre-sterilized soluble inorganic nutrients (phosphorus at 10-100 µM NaH₂PO₄, iron at 1-20 µM FeCl₃ with equimolar EDTA as a chelator, molybdenum at 0.1-10 µM Na₂MoO₄·2H₂O) and organic nutrients (mannitol at 1-10 mM, maltose at 1-10 mM) were added either individually or combined, using sterile plasticware for dispensing purposes. Particulate matter was derived from natural organic detrital sources in coastal Atlantic waters, including homogenized, rinsed (in sample water), sterilized suspensions of the common marshgrass *Spartina alterniflora* and seagrass *Zostera marina*. Because such particle sources provided both colonizable surfaces and leached organic/inorganic nutrients, we also examined the effects of chemically inert surfaces on microbial N₂ fixing potentials. These included cleaned, combusted glass-fibre filters (homogenized Whatman GF/C filters) and cleaned 20-µm mesh nylon screens suspended in bioassay vessels.

Nitrogen fixation responses to various treatments were assessed as nitrogenase activity measurements, using the acetylene reduction assay¹¹. Parallel measurements of dissolved O₂ concentration ([O₂]) in the bulk water phase and in gradients associated with surface-associated microzones were made with polarographic O₂-sensitive microelectrodes (tip diameter 5-10 µm), using a method described in detail previously^{12,13}.

Table 1 Effects of inorganic and organic particle and nutrient additions on coastal Atlantic Ocean N₂ fixation (nitrogenase activity) potentials

Treatment	Nitrogenase activity (nmol C ₂ H ₄ ml ⁻¹ h ⁻¹)
Control (No additions)	N.D.
0.1 g l ⁻¹ <i>Spartina</i> particles	1.2 ± 0.02
0.1 g l ⁻¹ <i>Zostera</i> particles	1.6 ± 0.04
5.0 mg l ⁻¹ mannitol	1.4 ± 0.03
0.1 g l ⁻¹ <i>Spartina</i> + 5.0 mg l ⁻¹ mannitol	3.6 ± 0.08
0.1 g l ⁻¹ glass fibre filters	0.6 ± 0.02
0.1 g l ⁻¹ glass fibre filters + 5.0 mg l ⁻¹ mannitol	2.5 ± 0.04
0.1 g l ⁻¹ Nylon screening	1.1 ± 0.03
0.1 g l ⁻¹ Nylon screening + 5.0 mg l ⁻¹ mannitol	3.1 ± 0.09
10-50 µM P	N.D.
1-10 µM Mo	N.D.
1-20 µM Fe	N.D.
P* + Mo* + Fe*	N.D.
0.1 g l ⁻¹ <i>Spartina</i> + P* + Mo* + Fe*	1.1 ± 0.03
0.1 g l ⁻¹ Nylon screening + P* + Mo* + Fe*	0.9 ± 0.02

Sterilized 1-l Nalgene transparent polycarbonate Erlenmeyer flasks fitted with foam plugs and magnetic stirring bars were used as incubation vessels. Incubations were conducted under intermittent (12 h light: 12 h dark) illumination (200 µEinst m⁻² s⁻¹ photosynthetically active radiation; cool white and Grolux fluorescent light). Incubations were maintained at sampling temperatures. All individual and combined nutrient/particle/screening additions were conducted in triplicate. Acetylene reduction assays in this and subsequent experiments were conducted on 20-ml subsamples incubated in stoppered 30-ml serum bottles. In the case of screening, a proportional (to original sample volume) piece of screening was placed in flasks. Standard errors of triplicated treatments are shown. N.D., nitrogenase activity not detected. For the gas chromatograph used (Shimadzu GC-9A coupled to a CR3 reporting integrator), the lower limit of nitrogenase activity detection is 0.05 nmol C₂H₄ ml⁻¹ h⁻¹.

* A range of individual P, Mo and Fe concentrations was employed in combined additions.

Regardless of season, additions of sterile *Spartina* or *Zostera* particles as well as nylon screening elicited N₂ fixation in these waters (Table 1). Significant (>3 times background ethylene content of the Matheson Inc. ultra-high purity grade acetylene used) rates of C₂H₄ production were detected within 24 h of particle/screen additions. Sterile glass-fibre filter particles alone often failed to elicit nitrogenase activity; however, combined with dissolved organic matter (mannitol or maltose), nitrogenase activity was readily observed within 36 h. Addition of dissolved organic matter alone also led to the development of nitrogenase activity; development of nitrogenase activity was linked to the formation of detrital flocs, resembling previously-described 'marine snow'¹⁴.

Maximum rates of nitrogenase activity were obtained when organic and inorganic particles were added together with dissolved organic matter (Table 1). Additions of phosphorus, iron and molybdenum, either alone or combined, over a range of concentrations, did not elicit nitrogenase activity (Table 1). In combination with particles, inorganic nutrient additions failed to further enhance nitrogenase activity.

To determine whether the nitrogenase activities reported in Table 1 were significant in supporting particle and screen-associated bacterial growth, we estimated rates of N₂ fixation based on the theoretical ratio of 3:1 for acetylene reduction to N₂ fixation^{5,6,7}. We obtained a mean N₂ fixation rate of 0.5 nmol N₂ ml⁻¹ h⁻¹ or 1.4 × 10⁻⁸ g N ml⁻¹ h⁻¹. Acridine orange-based bacterial counts in samples containing *Spartina* and *Zostera* detritus or nylon screens ranged from 2-5 × 10⁶ cells ml⁻¹ (both attached and free-floating cells). Average cell size was 0.1 µm³ (ref. 15). Using Bratback's¹⁶ volumetric value for mean bacterial cell carbon content (5.6 × 10⁻¹³ g C µm⁻³),

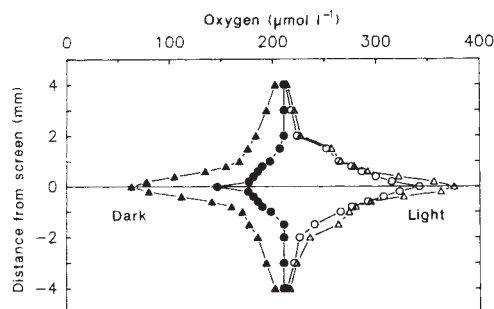


Fig. 1 Microprofiles of oxygen concentration $[O_2]$ near microbially colonized nylon screens. Sterile screens were incubated in sterilized, clear 6-l polycarbonate jars containing freshly sampled seawater. Jars were fitted with magnetic stirrers and clear lids. Circles indicate the distribution of O_2 early in the colonization phase, microbial film ~ 200 – $300 \mu\text{m}$ thick. Triangles show the distribution when the film on the screen is $\sim 1 \text{ mm}$ thick. Solid symbols, $[O_2]$ during darkness, open symbols, during illumination with $\sim 100 \mu\text{Einstein m}^{-2} \text{ s}^{-1}$.

we derived a mean cell carbon content of $5.6 \times 10^{-14} \text{ g C cell}^{-1}$. Previously obtained C:N ratios for coastal bacteria range from 10 to 12 (ref. 6). The cellular N content of our samplings thus averages $5.6 \times 10^{-15} \text{ g N cell}^{-1}$; volumetrically, the biomass is of the order of $2 \times 10^{-8} \text{ g N ml}^{-1}$. Considering an average N_2 fixation rate of $1.4 \times 10^{-8} \text{ g N ml}^{-1} \text{ h}^{-1}$, enough fixed nitrogen is provided to allow bacterial biomass to double every 1.4 h. Accordingly, the acetylene reduction rates given in Table 1 appear significant and adequate to sustain growth of these bacterial communities.

In all cases detectable rates of nitrogenase activity coincided with surficial, microzone-oriented O_2 depletion as measured by microelectrodes (Fig. 1). Recent work has shown localized O_2 depletion to be a common feature of a variety of detrital aggregates and flocs suspended in seawater^{12,17}. Of interest here were $[O_2]$ gradients associated with surfaces of nylon screens suspended in 6-l sterilized, clear polycarbonate jars. Screens supported extensive bacterial (within 24 h) and cyanobacterial (after 36 h) colonization. Screen-associated $[O_2]$ gradients and nitrogenase activity were detected during early (largely bacterial) and later (bacterial and cyanobacterial) colonization periods (Fig. 1). Supplementation with either mannitol or maltose led to stronger $[O_2]$ gradients and lower $[O_2]$ values close to the screen surface. Lower overall O_2 concentrations were consistently accompanied by enhanced nitrogenase activity. Inorganic nutrient (P, Fe, Mo) supplementation failed to stimulate either $[O_2]$ gradient formation or nitrogenase activity above controls (no nutrient additions) or organic matter additions.

For the first 36 h no significant differences in nitrogenase activity were observed between continuously illuminated and dark conditions (Fig. 2a). After 36 h noticeable colonization by the coccoid cyanobacterium *Synechocystis* sp. took place under illumination. Although *Synechocystis* sp. was readily observed on screens, it remained virtually undetectable (by microscopic counts) in the surrounding bulk-water phase. Once *Synechocystis* sp. established itself as a dominant screen epiphyte, illumination patterns became crucial in dictating screen-associated N_2 fixation potentials. Continuous illumination led to some enhancement of nitrogenase activity compared to permanently-dark conditions (Fig. 2a). After seven days of continuous illumination, however, *Synechocystis* sp. and bacterial biomass attached to screens notably declined, with previous growth falling off the screens. In contrast, intermittent (12 h on; 12 h off) illumination led to maximum nitrogenase activity (Fig. 2a). Vigorously growing colonies of *Synechocystis* sp. coated the screens within a week under intermittent illumination. Whereas illuminated periods produced supersaturated O_2 concentrations

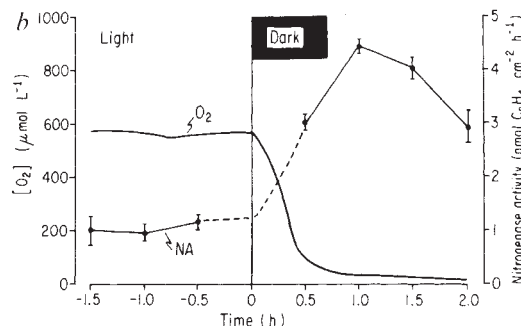
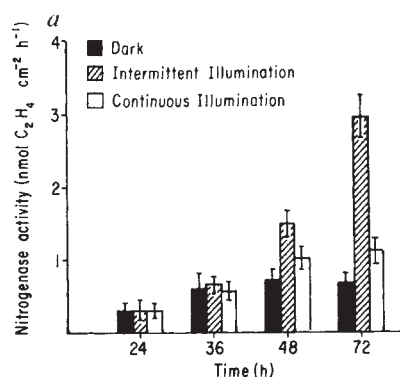


Fig. 2 a, Nitrogenase activity (NA) associated with suspended (in 6-l polycarbonate jars) nylon screens held under either continuous darkness, intermittent illumination or continuous illumination. Screens were incubated at seawater sampling temperatures. Three-hour NA determinations were conducted every 24 h after suspension of screens in seawater. Variability (three determinations) is shown with error bars. After 36 h, intermittent as well as continuously illuminated screens supported colonies of the cyanobacterium *Synechocystis* sp. b, Parallel microelectrode determinations of $[O_2]$ and nitrogenase activity (NA) associated with a nylon screen incubated for 7 days in seawater. Screens were incubated at seawater sampling temperature. The screen was heavily colonized by a mixed microbial community consisting of a variety of filamentous and rod-shaped eubacteria and the cyanobacterium *Synechocystis* sp. Relative changes of O_2 concentration and NA during the transition from illuminated ($200 \mu\text{Einstein m}^{-2} \text{ s}^{-1}$ photosynthetically active radiation, to dark conditions are shown. The attached microbial community was continuously monitored for $[O_2]$; NA determinations were made over 20-min intervals every 30 min. Dashed line, extrapolation based on data points closest to the transition from illuminated to dark conditions. Error bars, observed ranges for triplicate NA determinations.

on screen surfaces, dark periods quickly (within 30 min) led to extremely low O_2 concentrations in identical regions (Fig. 2b). Maximum rates of screen-associated nitrogenase activity were consistently observed within the first 1–3 h of dark periods (Fig. 2b). It is worth noting that nitrogenase activity was not detectable in the bulk phase during these experiments. Agitation (by shaking) of screens led to immediate decreases in surface-associated nitrogenase activity potentials under either illuminated or dark conditions (Fig. 3). These effects are presumably due to the contemporaneous enhancement of O_2 delivery to the cells during shaking.

These results illustrate that surface associated microzones enriched with readily used (by heterotrophic microorganisms) sources of organic matter can support an N_2 -fixing microbial community in nitrogen-depleted, oxic Atlantic coastal waters. At any time of the year, these waters appear to contain sufficient numbers of microorganisms potentially capable of N_2 fixation in the presence of O_2 -depleted microzones. It remains to be shown whether such microorganisms, when present as free-

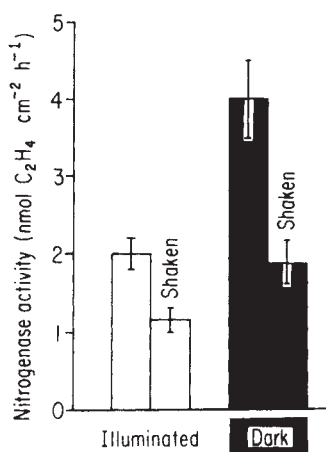


Fig. 3 Nitrogenase activity determined on illuminated compared to dark-incubated nylon screening in 1-l polycarbonate flasks exposed to unshaken or shaken (500 r.p.m. on an orbital shaker) conditions. Nitrogenase activity was assayed during a 3-h incubation period. Both dark and illuminated screens had previously been grown for seven days under an intermittent light (200 $\mu\text{Einsteins m}^{-2} \text{s}^{-1}$ photosynthetically active radiation) regime. Error bars, observed ranges of triplicate NA determinations.

floating populations, contain fully synthesized but inactivated (by O_2) nitrogenase or whether synthesis takes place only when O_2 -depleted microzones are encountered. In either case, organic matter necessary for both the creation of O_2 -depleted microzones and as an energy source supporting N_2 fixers can either be supplied exogenously or generated endogenously by photoautotrophic microorganisms growing together with N_2 fixers. Intermittent light/dark periods (as opposed to continuous light) are required to obtain maximum nitrogenase activity. This requirement is linked to the fact that under continuous illumination uninterrupted, photosynthetically generated O_2 accumulation in these consortia leads to the destruction of O_2 -depleted microzones, thus promoting persistent O_2 inhibition of nitrogenase activity. We are now examining the relative importance of *Synechocystis* sp., *Synechococcus* sp.¹⁸, other non-heterocystous cyanobacteria, and heterotrophic bacterial members of this consortium as potential contributors to microzone-mediated N_2 fixation.

This research was supported in part by the NSF, the North Carolina Sea Grant and the North Carolina Biotechnology Center. We thank B. Bebout, K. Crocker, V. Page, H. Page, J. Garner and L. Prufert for technical assistance.

Received 2 November 1987; accepted 15 January 1988.

- Dugdale, R. C. *Limnol. Oceanogr.* **10**, 67-73 (1967).
- Ryther, J. H. & Dunstan, W. M. *Science* **171**, 1008-1013 (1971).
- McCarthy, J. J. & Carpenter, E. J. in *Nitrogen in the Marine Environment* (eds Carpenter, E. J. & Capone, D. G.) 487-512 (Academic, New York, 1983).
- Carpenter, E. J. in *Nitrogen in the Marine Environment* (eds Carpenter, E. J. & Capone, D. G.) 65-104 (Academic, New York, 1983).
- Carpenter, E. J. & McCarthy, J. J. *Limnol. Oceanogr.* **20**, 389-401 (1975).
- Paele, H. W., Webb, K. L. & Weibe, W. J. in *The Ecology of Nitrogen Fixation* (ed. Broughton, W. J.) 193-241 (Oxford University Press, Oxford, 1981).
- Postgate, F. R. S. *The Fundamentals of Nitrogen Fixation* (Cambridge University Press, Cambridge, 1982).
- Williams, R. B. in *Proceedings of the Coastal Marsh and Estuary Management Symposium* (ed. Chabreck, R. H.) 185-201 (Louisiana State Univ. Div. of Continuing Education, Baton Rouge, 1972).
- Thayer, G. W. *Oecologia* **14**, 75-92 (1974).
- Paele, H. W. *Nature* **316**, 474-479 (1985).
- Stewart, W. D. P., Fitzgerald, C. R. & Burris, R. H. *Proc. natn. Acad. Sci. U.S.A.* **58**, 2071-2078 (1967).
- Paele, H. W. & Prufert, L. E. *Appl. Envir. Microbiol.* **53**, 1078-1087 (1987).
- Carlton, R. G. & Wetzel, R. G. *Can. J. Bot.* **65**, 1031-1037 (1987).
- Allredge, A. L. & Cox, J. L. *J. Mar. Res.* **40**, 517-521 (1984).
- Palumbo, A. V., Ferguson, R. L. & Rublee, P. A. *Appl. Envir. Microbiol.* **48**, 157-164 (1984).
- Bratbak, G. *Appl. Envir. Microbiol.* **49**, 1488-1493 (1985).
- Allredge, A. L. & Cohen, Y. *Science* **235**, 689-691 (1987).
- Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H. & Arai, T. *Nature* **323**, 720-722 (1986).

pH-sensitive activation of the intracellular-pH regulation system in squid axons by ATP- γ -S

W. F. Boron*, E. Hogan* & J. M. Russell†

* Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

† Department of Physiology and Biophysics, University of Texas Medical Branch, Galveston, Texas 77550, USA

The regulation of intracellular pH (pH_i) is essential for normal cell function¹, and controlled changes in pH_i may play a central role in cell activation². Sodium-dependent $\text{Cl}^-/\text{HCO}_3^-$ exchange is the dominant mechanism of pH_i regulation in the invertebrate cells examined³⁻⁶, and also occurs in mammalian cells^{7,8}. The transporter extrudes acid from the cell by exchanging extracellular Na^+ and HCO_3^- (ref. 9) (or a related species) for intracellular Cl^- (refs 3, 4). It is blocked by the stilbene derivatives DIDS (4,4'-diisothiocyano-stilbene-2,2'-disulphonate, ref. 10) and SITS (4-acetamido-4'-isothiocyano-stilbene-2,2'-disulphonate, ref. 3), and has a stoichiometry of two intracellular H^+ neutralized for each Na^+ taken up and each Cl^- extruded by the axon¹¹. Because the inwardly-directed Na^+ concentration gradient is sufficiently large to energize both the HCO_3^- influx and Cl^- efflux, this electroneutral exchanger could be a classic secondary active transporter, thermodynamically independent of ATP hydrolysis. However, at least in the squid axon, the exchanger has an absolute requirement for ATP (ref. 3). Thus, a major unresolved issue is whether this Na-dependent $\text{Cl}^-/\text{HCO}_3^-$ exchanger stoichiometrically hydrolyses ATP (the pump hypothesis), or whether ATP activates the transporter by a mechanism such as phosphorylation or simple binding (the activation hypothesis). We have now explored the role of ATP in pH_i regulation by dialysing axons with the ATP analogue ATP- γ -S. In many systems, ATP- γ -S is an acceptable substrate for protein kinases^{12,13}, whereas the resulting thiophosphorylated proteins are not as readily hydrolysed by phosphatases as are phosphorylated proteins^{14,15}. Our results rule out the pump hypothesis, and show that the basis of the axon's ATP requirement is the pH-dependent activation (by, for instance, phosphorylation or ATP binding) of the exchanger itself, or of an essential activator.

The experiments were performed on single squid giant axons. The intracellular milieu was controlled by internal dialysis¹⁶, and pH_i was monitored with a pH-sensitive glass microelectrode^{17,18}. Because acid-extruding transporters such as the Na-dependent $\text{Cl}^-/\text{HCO}_3^-$ exchanger are inactive at alkaline pH_i but activated at low pH_i , they are best studied by acutely acid-loading the cell and monitoring the subsequent pH_i recovery, which is mediated by an acid-extruding mechanism. In our experiments, axons were acid-loaded by dialysing with a low-pH dialysis fluid (DF).

Earlier experiments on squid axons had shown that acid extrusion is reversibly blocked by cyanide⁹, but is restored when cyanide-treated axons are dialysed with ATP (ref. 3). Figure 1A illustrates the results of an experiment in which an axon was dialysed with DF at pH 6.7, containing no ATP but ATP- γ -S at a nominal concentration of 8 mM. The extracellular solution (artificial seawater, ASW) was initially HCO_3^- -free, and throughout the experiment contained 1 mM cyanide. Dialysing with the low-pH DF caused the pH_i to decline slowly (segment a-b on the trace in Fig. 1). When dialysis was halted (point b), returning control of pH_i to the axon, the pH_i failed to recover (b-c) as the Na-dependent $\text{Cl}^-/\text{HCO}_3^-$ exchanger has an absolute requirement for HCO_3^- . Addition of CO_2 and HCO_3^- to the ASW caused a rapid pH_i fall (c-d) due to the CO_2 influx, followed by a slower pH_i recovery (d-e) presumably due to acid extrusion

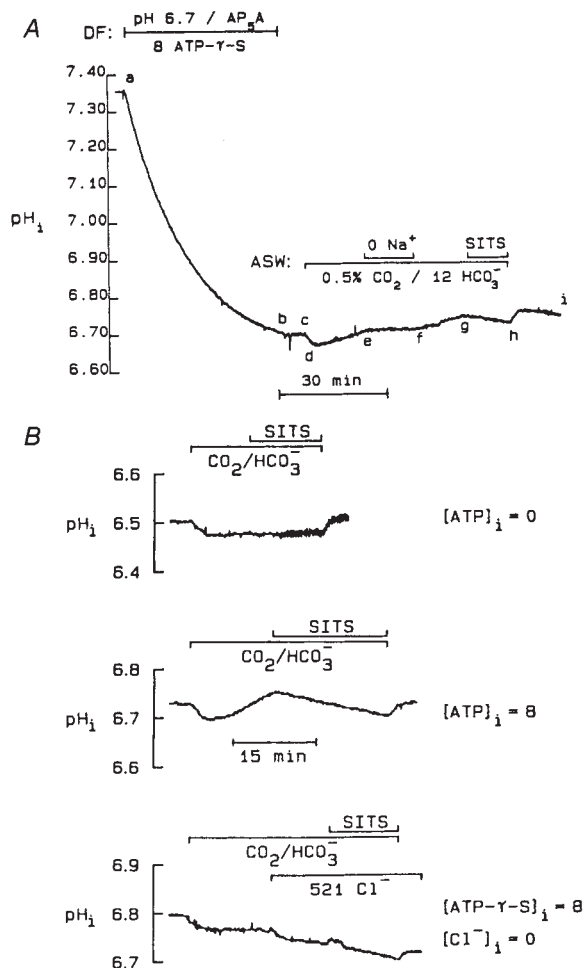


Fig. 1 *A*, Support of acid extrusion by ATP- γ -S. Between points a and b, the axon was dialysed with a fluid at pH 6.7 containing 8 mM ATP- γ -S. The complete composition is given below. The artificial sea water (ASW) to which the axon was exposed was nominally HCO_3^- -free except as indicated, and contained 1 mM cyanide (CN^-). After dialysis was halted, application of HCO_3^- produced a pH_i recovery (d-e) that was blocked by replacing Na^+ with Li^+ (e-f), or by applying 0.5 mM SITS (g-h). *B*, Control experiments, performed with a protocol similar to that described above. In the top record, the axon had been dialysed for 60 min with a solution at pH 6.5 containing neither ATP nor ATP- γ -S. There was no SITS-sensitive pH_i recovery. In the middle record, the axon had been dialysed for 60 min with a fluid at pH 6.6 containing 8 mM ATP. There was substantial SITS-sensitive pH_i recovery. In the bottom record, the axon had been dialysed for 60 min with a fluid at pH 6.6 containing 8 mM ATP- γ -S, but no Cl^- . The ASW was also Cl^- -free. Addition of HCO_3^- elicited no pH_i recovery, even after Cl^- was returned to the ASW, and SITS was without effect.

Methods. Unless otherwise stated, the dialysis fluid had the following composition (in mM): 0 Na^+ , 413 K^+ , 7 Mg^{++} , 400 Cl^- , 14 glutamate, 20 MES, 0.5 phenol red, and 8 ATP- γ -S (Boehringer-Mannheim). Because the ATP- γ -S is reported to contain about 15% ADP, we added 1.25 mM diadenosine pentaphosphate (AP_5A) in some experiments to block the conversion of ADP to ATP. We found that omission of AP_5A had no effect on the results. The standard HCO_3^- -free ASW had the following composition (in mM): 425 Na^+ , 10 K^+ , 3 Ca^{2+} , 62.5 Mg^{2+} , 561 Cl^- , 10 EPPS, 0.1 EDTA, 1 CN^- . The pH was 8.00. In the HCO_3^- -containing ASW, 12 mM KHCO_3 replaced 10 mM KCl , [EPPS] was raised to 30 mM with corresponding reductions in $[\text{Mg}^{2+}]$ and $[\text{Cl}^-]$ to keep osmolarity constant, and the solutions were equilibrated with 0.5% CO_2 , 99.5% air. J_H was calculated from the measured rate of pH_i change, the measured axon diameter, the intrinsic intracellular buffering power (assumed equal to the buffering power of the DF), and the CO_2 buffering power (computed from P_{CO_2} and the measured pH_i).

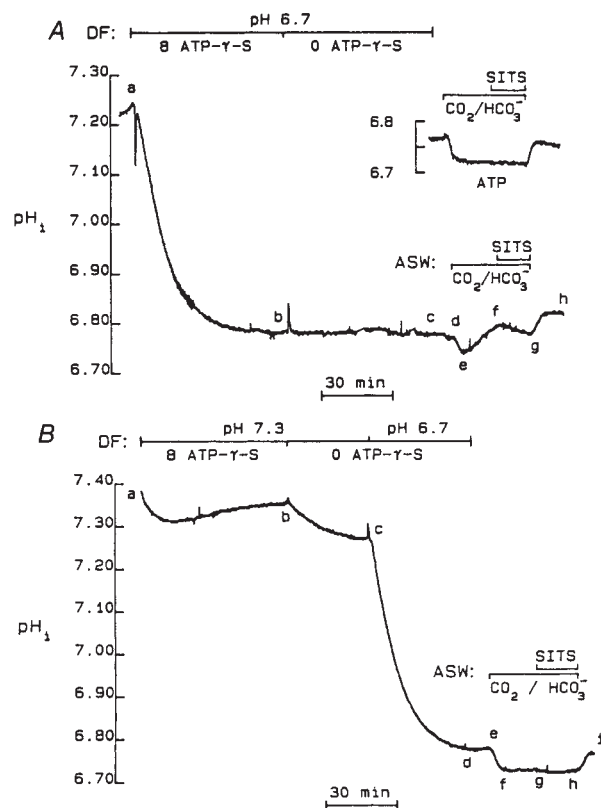


Fig. 2 Effect on pH_i recovery of applying and then washing out ATP- γ -S. *A*, Application of ATP- γ -S at low pH_i . The axon was dialysed with ATP- γ -S at pH 6.7 for 60 min, and then dialysed with a nucleotide-free fluid for an additional 60 min. There was a substantial pH_i recovery (e-f) sensitive to SITS (f-g). The inset illustrates an experiment in which the 60-min nucleotide-free washout was preceded by a 60-min period of dialysis with 8 mM ATP. There was virtually no SITS-sensitive pH_i recovery. *B*, Application of ATP- γ -S at high pH_i . The protocol was identical to that described above, except that the ATP- γ -S was applied in a DF at pH ~ 7.35 (a-b), and then washed out first in a DF at pH ~ 7.28 (b-c), and then in one at pH 6.7. After a total nucleotide-free washout of 60 min, there was negligible SITS-sensitive pH_i recovery from the acid load.

by the Na-dependent Cl-HCO_3 exchanger. The pH_i recovery was blocked by removal of the external Na^+ (e-f, Na^+ replaced with Li^+), but resumed when the Na^+ was returned (f-g). The Na-dependent component of the pH_i recovery corresponded to a net acid-extrusion rate (that is d-e versus e-f) of $11.6 \pm 2.7 \text{ pmol cm}^{-2} \text{ s}^{-1}$ (averaged over experiments, $n = 3$). Finally, application of SITS blocked the pH_i recovery, and revealed a small underlying acidification (g-h). The SITS-sensitive acid extrusion rate (J_H), the difference between acid-extrusion rates in the absence (f-g) and presence of SITS (g-h), was $13.7 \text{ pmol cm}^{-2} \text{ s}^{-1}$ for this axon, and averaged $12.4 \pm 1.2 \text{ pmol cm}^{-2} \text{ s}^{-1}$ in eight experiments. As shown by the top record of Fig. 1*B* (in which the portion comparable to segment a-b in Fig. 1*A* is omitted), there is no HCO_3^- -dependent/SITS-sensitive pH_i recovery when the axon is dialysed for 60-90 min with a solution containing no ATP or ATP- γ -S (mean $J_H = 0.9 \pm 1.2 \text{ pmol cm}^{-2} \text{ s}^{-1}$, $n = 3$). On the other hand, as can be seen in the second record of Fig. 1*B*, the pH_i recovery that occurs after dialysis with 8 mM ATP ($J_H = 12.9 \pm 2.4 \text{ pmol cm}^{-2} \text{ s}^{-1}$, $n = 3$) is comparable to that observed in axons dialysed with ATP- γ -S.

The Na-dependent Cl-HCO_3 exchanger has an absolute requirement for intracellular Cl^- (refs 3, 11). Thus, if the acid-extrusion activity supported by ATP- γ -S is mediated by the Na-dependent Cl-HCO_3 exchanger, it should be blocked by removal of internal Cl^- . As shown by the bottom record of Fig.

1B, ATP- γ -S does not support acid extrusion in an axon dialysed for 60 min with a Cl-free solution (mean $J_H = 1.5 \pm 1.3$, $n = 5$). Thus, we have shown that the pH_i recovery supported by ATP- γ -S has the characteristics of Na-dependent Cl-HCO₃ exchange: a requirement for external HCO₃⁻, external Na⁺ and internal Cl⁻, and sensitivity to SITS.

Because ATP- γ -S does not support such E₁-E₂ ATPases as the Ca²⁺ pump^{19,21} and the Na-K pump²¹, it is unlikely that it is acting via an ATPase in the present experiments. Moreover, if ATP- γ -S were serving as a fuel for an ATPase, acid extrusion presumably would depend on the continued presence of ATP- γ -S. We directly tested this possibility in the experiment of Fig. 2A. This axon was dialysed for 1 h with DF at pH 6.7 containing 8 mM ATP- γ -S (a-b), and then for 1 h with DF at the same pH, but lacking ATP- γ -S (b-c). After dialysis was halted (c), the application of CO₂ and HCO₃⁻ (d) produced a pH_i recovery (e-f) that was blocked by SITS (f-g, mean $J_H = 13.2 \pm 1.8$ pmol cm⁻² s⁻¹, $n = 6$), even though the ATP- γ -S had been washed out. As shown in the inset to Fig. 2A, if the axon is acid-loaded by dialysing for 1 h with DF containing ATP instead of ATP- γ -S, J_H is nearly zero after washing out the ATP ($J_H = 1.1 \pm 0.4$ pmol cm⁻² s⁻¹, $n = 6$). We verified that washing out for 1 h is sufficient to reduce the concentration of ATP- γ -S or ATP in extruded axoplasm to less than 10 μ M. Unless the very low levels of ATP- γ -S that remain after washing out are sufficient to fuel a hypothetical pump, one must conclude that the continued acid extrusion in the nominal absence of ATP- γ -S is the result of either maintained thiophosphorylation of, or maintained binding of ATP- γ -S to, the exchanger or an essential activator.

In the above experiments, the ATP- γ -S was always applied and washed out at low pH_i . In the experiment of Fig. 2B, however, the 8 mM ATP- γ -S was applied at pH 7.3 (a-b). The ATP- γ -S was then washed out by dialysing with ATP- γ -S-free solutions, first at pH 7.3 (b-c), and then at pH 6.7 (c-d). After dialysis was halted (d), the addition of CO₂ and HCO₃⁻ (e) elicited no pH_i recovery (f-g), and SITS (g-h) was without effect (mean $J_H = 1.1 \pm 0.9$ pmol cm⁻² s⁻¹, $n = 6$). Thus, under alkaline conditions, either activation of transport by ATP- γ -S is inhibited and/or deactivation is enhanced. We tested the latter possibility by dialysing the axon with 8 mM ATP- γ -S at pH 6.7 for one hour (thereby activating the transporter, see Fig. 2A), and then washing out the ATP- γ -S for an additional hour. During the washing out of ATP- γ -S, however, the pH_i was raised above 7.3 for varying periods. As shown in Fig. 3A, raising the pH_i above 7.3 for only ~5 min produced a modest reduction in the SITS-sensitive pH_i recovery rate (g-h versus h-i; mean $J_H = 11.2 \pm 0.6$ pmol cm⁻² s⁻¹, $n = 3$). On the other hand, as shown in Fig. 3B, extending the period of alkaline ATP- γ -S washout to 1 h prevented any subsequent pH_i recovery (mean $J_H = 0.4 \pm 1.6$ pmol cm⁻² s⁻¹, $n = 3$). The data, summarized in Fig. 3C, imply that the ATP- γ -S-activated exchanger is slowly deactivated at alkaline pH_i . The most straightforward explanation for these data is that high pH_i enhances the deactivation of the transporter by a mechanism such as dethiophosphorylation or the dissociation of ATP- γ -S. The possibility of a second pH_i -sensitive effect (that is, the inhibition of activation at high pH_i) was not addressed in the present experiments.

The first major observation of this paper is that the Na-dependent Cl-HCO₃ exchanger is not a pump, but requires ATP either for the activation of the exchanger *per se* or of an essential activator. Given the previous work with ATP- γ -S, we feel that the most likely mechanism of activation in our experiments is thiophosphorylation, though we cannot rule out the possibility that ATP- γ -S activates the transporter by binding with an extremely high affinity. In principle, phosphorylation and binding could be distinguished by employing a non-hydrolysable ATP analogue, such as AMPPNP. However, such an experiment might not be definitive, inasmuch as we found that acid extrusion was not supported in five axons dialysed with the hydrolysable analogue AMP CH₂PP, at concentrations of 5 to 20 mM (data

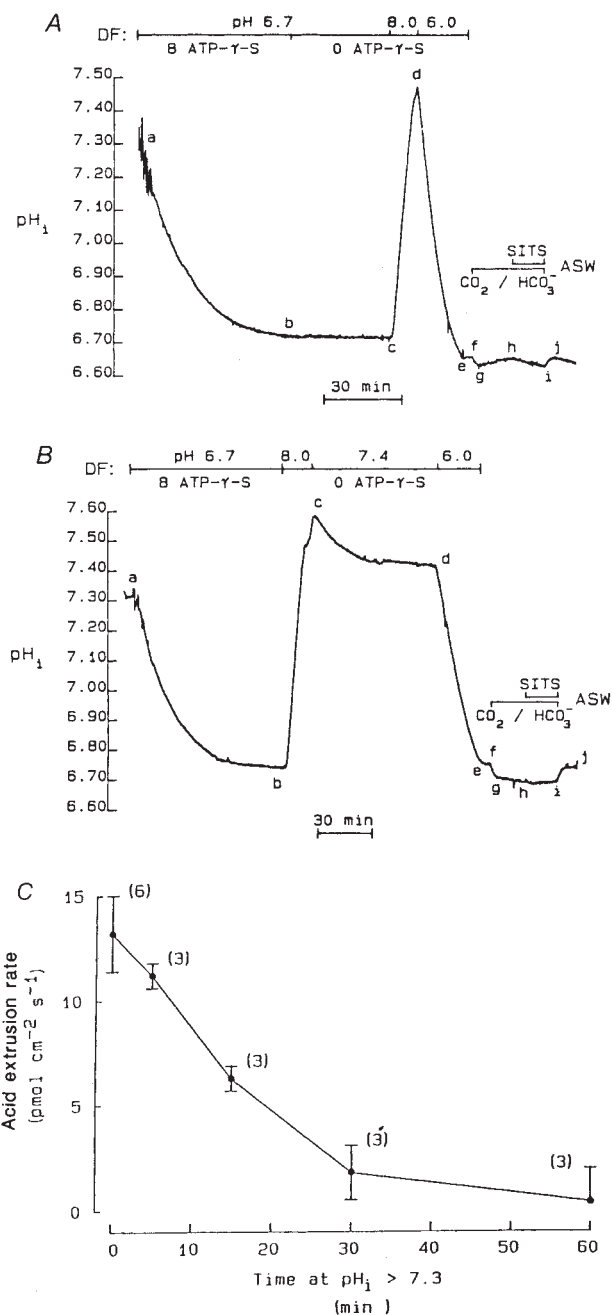


Fig. 3 Time-dependence of the inactivation of acid extrusion by alkaline pH_i . **A**, The axon was treated with ATP- γ -S for 60 min at pH_i 6.7, and then dialysed with a nucleotide-free fluid for an additional 60 min. For about 5 min, the pH_i was elevated above 7.3. Upon reacidification, there was still a substantial SITS-sensitive pH_i recovery. **B**, After the initial exposure to ATP- γ -S at low pH_i , the axon was held at $pH_i > 7.3$ for 60 min during the nucleotide-free washout. There was virtually no SITS-sensitive pH_i recovery following an acid load. **C**, Summary of 18 similar experiments. During the 60-min nucleotide-free washout, pH_i was elevated above 7.3 for periods ranging from 0 to 60 min.

not shown). The second and perhaps more intriguing observation is that, at least when ATP- γ -S is employed, the state of activation of this pH_i regulator depends critically on the parameter being regulated (that is, pH_i). To the extent that the interaction with the exchanger is similar for ATP and ATP- γ -S, it is likely that a portion of the pH_i dependence that characterizes acid extrusion in the squid axon is attributable to the pH_i

dependence of the ATP-dependent activation/deactivation mechanism.

We thank Ms Emily Tyner for her assistance in making the drawings. Supported by grants from the NIH.

Received 19 October 1987; accepted 8 January 1988.

1. Roos, A. & Boron, W. F. *Physiol. Rev.* **61**, 296-434 (1981).
2. Moolenaar, W. H. A. *Rev. Physiol.* **48**, 363-376 (1986).
3. Russell, J. M. & Boron, W. F. *Nature* **264**, 73-74 (1976).
4. Thomas, R. C. *J. Physiol., Lond.* **273**, 317-338 (1977).
5. Boron, W. F., McCormick, W. C. & Roos, A. *Am. J. Physiol.* **240**, C80-C89 (1981).
6. Moody, W. J. Jr *J. Physiol., Lond.* **316**, 293-308 (1981).
7. L'Allemain, G., Paris, S. & Pouyssegur, J. *J. biol. Chem.* **260**, 4877-4883 (1985).
8. Rothenberg, P., Glaser, L., Schlesinger, P. & Cassel, D. *J. biol. Chem.* **258**, 12644-12653 (1983).
9. Boron, W. F. & De Weer, P. *Nature* **259**, 240-241 (1976).
10. Thomas, R. C. *Nature* **262**, 54-55 (1976).
11. Boron, W. F. & Russell, J. M. *J. gen. Physiol.* **81**, 373-399 (1983).
12. Gratecos, D. & Fischer, E. H. *Biochem. biophys. Res. Commun.* **58**, 960-967 (1974).
13. Sun, I. Y., Johnson, E. M. & Allfrey, V. G. *J. biol. Chem.* **255**, 742-747 (1980).
14. Hauptmann, M., Winson, D. F. & Erecinska, M. *Biochem. Pharmacol.* **8**, 1247-1254 (1985).
15. Brooks, J. C. & Brooks, M. *Life Sci.* **37**, 1869-1875 (1985).
16. Brinley, F. J. Jr & Mullins, L. J. *J. gen. Physiol.* **50**, 2303-2331 (1967).
17. Hinkle, J. A. M. in *Glass Electrodes for Hydrogen and Other Cations. Principle and Practice*. (ed. Eisenman, G.) 464-477 (Dekker, New York, 1967).
18. Boron, W. F. & De Weer, P. *J. gen. Physiol.* **67**, 91-112 (1976).
19. Yates, D. W. & Duance, V. C. *Biochem. J.* **159**, 719-728 (1976).
20. Yasuoka, K., Kawakita, M. & Kazi, Y. *J. biochem., Tokyo* **91**, 1629-1637 (1982).
21. DiPolo, R. & Beauge, L. *Biochem. biophys. acta* **897**, 347-354 (1987).

Identification and characterization of *E. coli* type-1 pilus tip adhesion protein

Mark S. Hanson* & Charles C. Brinton Jr†

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

The type-1 pilus of *Escherichia coli* is the prototype of this class of hair-like, multimeric adhesive organelles^{1,2}. This pilus mediates adherence to mannose-containing receptors³ on mucosal epithelia and other cells⁴. The type-1 pilus, in one of several serological variants, is expressed by nearly all *E. coli* strains^{5,6}, and its promotion of colonization by pathogenic bacteria and the protective effects of purified pilus vaccines suggest that it is important as a bacterial virulence factor^{5,7,8}. Both the adhesive function and the serological variation of the type-1 pilus have been attributed to the thousand or so pilin protein monomers making up the pilus rods. This idea has been contradicted by our earlier observations on an *E. coli* strain expressing adhesion-defective pilin⁹. More recent genetic evidence also indicates that auxiliary pilus proteins are required for adhesive function^{10,11}. We report here the identification of three previously undetected integral minor proteins on the type-1 pilus, and show that one of them is the receptor-binding adhesin. This protein is antigenically conserved among strains with different pilin serotypes and is located at the pilus tip.

Escherichia coli B. strain American (Bam, ref. 2), does not express flagella, sex pili, or pili other than type 1 under any growth conditions so far examined⁶. Pili were purified in high yield from this strain² to apparent homogeneity, as assessed by SDS-PAGE. SDS-PAGE and autoradiography of ¹²⁵I-labelled¹² pili revealed three heavily iodinated species (data not shown) of relative molecular mass (*M_r*) 28,000 (28K), 16.5K and 14.5K respectively, in addition to pilin with *M_r* 17K. A silver-staining protocol¹³ revealed these pilus-associated proteins on SDS gels resolving unlabelled pili (such as in Fig. 1a, lanes 1 and 2). A protein of *M_r* 36K that was loosely associated with the purified pili was removed using Triton X-100, but the remaining three species could not be removed from pili by ultracentrifugation

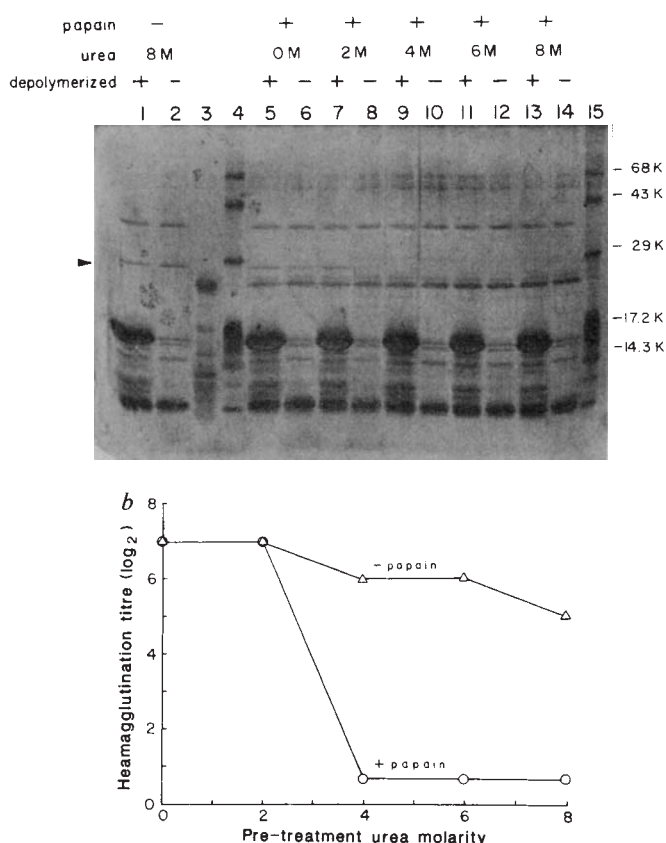


Fig. 1 Treatment of intact pili with papain in the presence of concentrated urea leads to selective degradation of the 28K minor protein and loss of HA activity. Pili ($200 \mu\text{g ml}^{-1}$) in various concentrations of urea were incubated in the presence or absence of papain ($8 \mu\text{g ml}^{-1}$) at 37°C for one hour. These mixtures were dialysed against distilled water and assayed for the extent of (a), proteolysis and (b) HA activity loss. Type-1 pilin subunit interactions are stable to SDS (ref. 21), requiring depolymerization treatment by boiling at $\text{pH} < 2$ (ref. 2) before SDS-PAGE²². a, Silver-stained SDS gel (15% acrylamide) showing both depolymerized pili (lanes 1, 5, 7, 9, 11, 13), and non-depolymerized pili (lanes 2, 6, 8, 10, 12, 14), which show the 16.5K and 14.5K bands more clearly. The 28K band is indicated by an arrow. Lane 3, papain marker; lanes 4 and 15, *M_r* standards. b, Twofold dilution series of papain/urea-treated pili were mixed with 2% guinea pig erythrocytes to determine relative HA activity by a titration end-point agglutination assay. Pilus HA activity was resistant to treatment with urea alone ($\blacktriangle$), but was abolished at urea concentrations of 4M and above when the pre-treatment included papain ($\circ$). Adding papain directly to the HA assay did not affect activity, indicating that the pilus erythrocyte receptor is resistant to papain under these conditions (not shown).

or by column chromatography, even in the presence of non-ionic detergents or concentrated urea or guanidine hydrochloride. Treatment of pili with SDS at elevated temperature (80 – 100°C) removed these three proteins and allowed their purification (M.S.H. and C.C.B., manuscript in preparation), but also inactivated the pilus binding function, as defined by haemagglutination (HA). Thus, the tightly bound 28K, 16.5K and 14.5K proteins are previously undetected integral minor proteins of the pilus. Densitometry of overloaded ($>30 \mu\text{g}$ pili) Coomassie blue-stained gels indicated that the ratio of 28K:17K (pilin):16.5K:14.5K proteins was $\sim 2:300\text{--}400:1:1$ (data not shown).

The denaturing conditions necessary for removal of the minor pilus proteins preclude determination of the HA activities of the resulting components. Type-1 pili are structurally and functionally resistant to proteolysis and to urea. Certain combinations of papain and concentrated urea were found to degrade

* Present address: University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA.
† To whom correspondence should be addressed.

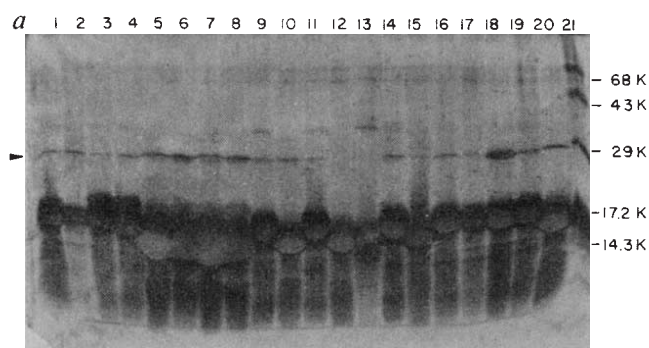
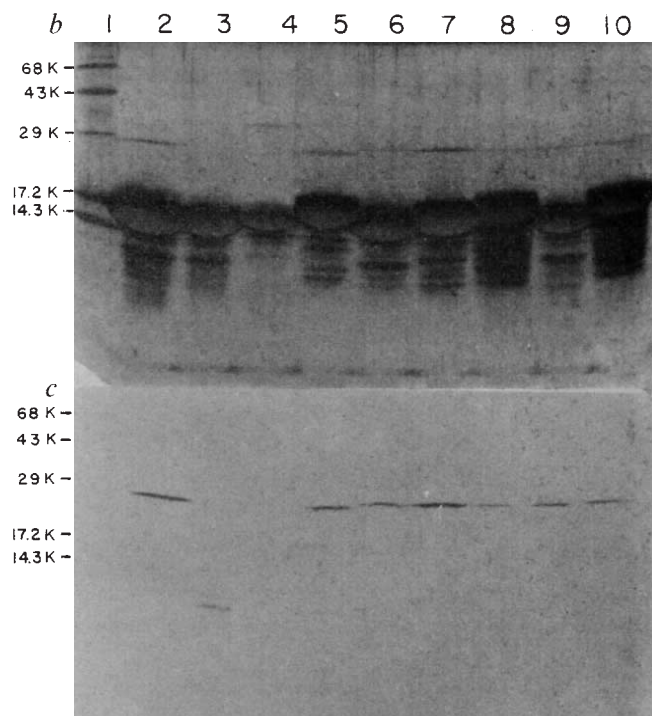


Fig. 2 The relative molecular mass (M_r) and serological cross-reactivity of the 28K protein is conserved among different type-1 pilus serotypes whose main pilin subunit is variable. *a*, Presence or absence of 28K protein in type-1 pili purified from *E. coli* isolated from different sources indicated by SDS-PAGE with silver-staining. These strains included enterotoxigenic isolates of bovine, porcine and human origin, as well as isolates from patients with pyelonephritis and asymptomatic bacteriuria. Pili from a cystitis isolate (*a*, lane 13; *b*, lane 4) were later found to be mannose-resistant. Pili from the bovine isolate 190 (lane 18) were contaminated with F41 pilin, M_r 29.5K (ref. 23), co-expressed with type-1 pili in this strain. Pili from the prototype strain Bam are in lane 20. Size markers are run in lane 21. The 28K band is indicated by the arrow. *b*, Serologically distinct type-1 pili from eight different *E. coli* strains (grown at 37 °C), plus mannose-resistant pili from cystitis strain C9, run on a 12.5% acrylamide SDS gel and silver-stained. *c*, A parallel gel transferred to nitrocellulose and immunoblotted; lane 1, M_r standards; lane 2, strain Bam (sewage isolate); lane 3, AW405 (laboratory strain); lane 4, C9 (human cystitis); lane 5, B9 (human enterotoxigenic); lane 6, H10407 (human enterotoxigenic); lane 7, RDEC/Nal-1 (rabbit enterotoxigenic); lane 8, B44 (bovine enterotoxigenic); lane 9, Br11 (human pyelonephritic); lane 10, 1459 (bovine enterotoxigenic). The band of low M_r in lane 3 is presumed to be a breakdown product of the 23K band observed when AW405 is grown at intermediate



temperatures (see Fig. 3). Electrophoretic transfer of proteins from the gel to nitrocellulose paper was as described²⁴. Before exposure to antibodies, the paper was blocked with 1% non-fat milk in 10 mM Tris, 0.85% NaCl, pH 7.4 (TBS). Mouse anti-Bam 28K protein (1:100 in [TBS plus] 1% milk) was incubated with the paper, washed five times in TBS, and reacted with horseradish peroxidase anti-mouse IgG conjugate in TBS+1% milk. After washing, bound antibody is revealed with 0.015% H_2O_2 plus the chromogen 3-amino-9-ethylcarbazole in 10 mM sodium acetate, pH 5.

only the 28K protein, leaving the other minor proteins and the pilus rod intact (Fig. 1*a*, lanes 9–14). After removal of the urea, pili treated with and without papain were assayed for HA activity. Those pili which had lost the 28K protein had also lost essentially all HA activity (Fig. 1*b*). Pili treated with concentrated urea alone or papain alone retained both the 28K protein and activity. These observations implied that the 28K protein was required for adhesion.

As type-1 pili from almost all the previously studied *E. coli* strains promote adhesion, we determined whether or not they all contained the 28K protein. Pili purified from the prototype strain Bam (Fig. 2*a*, lane 20) were compared with pili from the adhesion-deficient AW405 (ref. 9) (lane 12), and with pili from pathogenic *E. coli* strains of diverse clinical and geographical origin (lanes 1–11, 13–19). These comprised most of the known type-1 serotypes⁶. The 28K protein occurred ubiquitously among pili purified from these strains, and was more highly conserved in size than pilin (Fig. 2*a*). Pili purified from an *E. coli* strain from a cystitis patient showed a band at 34K rather than at 28K (lane 13); these pili promoted mannose-resistant HA and were therefore not type-1 pili. All pili with the 28K protein promoted strong HA (data not shown). The antigenic conservation of the 28K protein is addressed in Fig. 2*b* and *c* (see below).

The defect in expression of pilus HA activity in strain AW405 was partially relieved at lower growth temperatures. Pili were purified from AW405 grown at several temperatures and assayed for their minor protein content (Fig. 3*a* and *b*) and HA activity (Fig. 3*c*). At growth temperatures above 25 °C this strain expressed pili almost devoid of 28K protein. Some pili grown at intermediate temperatures had a 23K protein (Fig. 3*a*, 32 °C and 36 °C). This antigenically related species (Fig. 3*b*) seems to represent an inactive fragment of 28K. The observation that

lambda strains with certain amber mutations plated on AW405 at 30 °C, but not 42 °C, suggests that a Ts-nonsense suppressor may be responsible for expression of the 28K protein at lower temperatures (data not shown). The relative HA activities of the pili grown at different temperatures (Fig. 3*c*) correlated with their 28K-protein content at these temperatures. It is interesting that the amounts of 16.5K and 14.5K proteins on these pili were unaffected by growth temperature.

These results imply that the 16.5K and 14.5K proteins are insufficient to confer HA activity on the type-1 pilus, and that the 28K protein is the receptor-binding adhesion protein. We investigated this using antisera prepared against both purified 28K protein and purified pilin. Table 1 shows that pilin and the 28K protein were serologically unrelated by the enzyme-linked immunosorbent assay (ELISA). The level of antibodies against the 28K protein elicited by immunization with intact pili was below the level of sensitivity of this ELISA, but immunoblotting indicated that low levels of anti-28K protein were produced by whole pilus immunizations (data not shown). The HA-inhibitory activities of IgG antibody and corresponding Fab fragments prepared from anti-pilin and anti-28K protein sera are also shown in Table 1. IgG against both pilin and 28K protein inhibited HA, but on cleavage to monovalent Fab fragments, the antibodies against pilin lost their ability to inhibit HA. These results indicate that pilin-specific IgGs inhibited HA by agglutinating the pili, and that 28K protein-specific antibodies inhibited HA by blocking the pilus receptor-binding sites.

Anti-28K serum was used to show that the 28K protein is conserved not only in function and size, but also serologically. Serum raised against 28K protein purified from strain Bam (Fig. 2*c*, lane 2) reacts well with the 28K protein of type-1 pili from

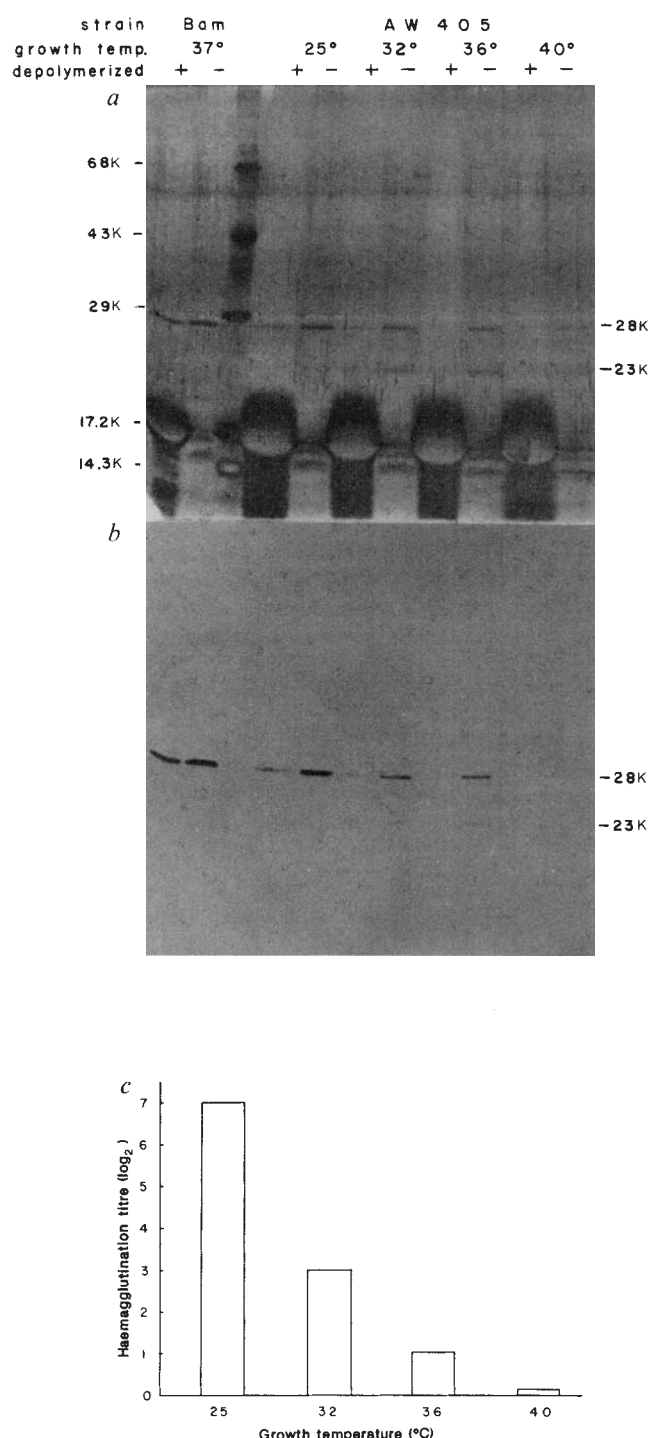


Fig. 3 Haemagglutination activity and the 28K pilus adhesion expression are coordinately expressed in a temperature-sensitive fashion in a pilus adhesion mutant. The adhesion-deficient pilated strain AW405 was grown at various temperatures for the production and purification of pili. *a*, And *b*, purified pili (15 μ g) from each of four growth temperatures were loaded on parallel SDS gels (12.5% acrylamide) in pairs of lanes for both depolymerized ($pH < 2$, 100 °C) and non-depolymerized samples. Similar samples of Bam pili (from 37 °C) were also included. *a*, Silver-stained gel and *b*, an immunoblot of a parallel gel with anti-28K protein serum. *c*, Haemagglutination activity of 1 mg ml⁻¹ samples of AW405 pili from four different temperatures were tested by diluting pilus crystal aggregates in 0.1 M MgCl₂ through a twofold series, mixing with 2% guinea pig erythrocytes and determining haemagglutination end-points. The graph shows the end-point activity titres of these pili. The decrease in 28K relative to pilin correlates with loss in HA activity.

HA-reactive pathogenic strains (lanes 5–10), but not with mannose-resistant pili from cystic strain C9 (lane 4).

We examined pili from adhesion-defective mutants (kindly provided by P. Orndorff) and find that at least two insertions in the type-1 pilus operon, which cause expression of HA-inactive pili¹¹, also lead to expression of pili lacking the 28K protein (data not shown). These same pili also have reduced amounts of the 16.5K and 14.5K proteins. We propose that *pilE* encodes the pilus adhesin as the product of this gene is similar in size to the 28K protein.

We have also considered the location of the 28K protein on the pilus. It has been observed that non-aggregated cell-free pili bind to erythrocytes but do not agglutinate them¹⁴, implying that pili are monovalent with respect to adhesion. Individual pili adsorbed to erythrocytes can bind anti-pilin, but not anti-28K protein antibodies. Longitudinal needle-like aggregates of pili ('crystals') were observed by dark-field microscopy to bind

Table 1 Antiserum cross-reactivity of type-1 pilus proteins and haemagglutination inhibition by rabbit antibodies against them

Protein	Antiserum		
	Rabbit anti-whole pili	Rabbit anti-pilin rod	Mouse anti-28K
Whole pili (1 μ g per well)	10,048	4,654	<20
Pilin rod (1 μ g per well)	8,832	4,573	<20
28K protein (0.1 μ g per well)	<20	<20	2,771
End-point concentrations (μ g ml ⁻¹)			
	For haemagglutination inhibition (HAI)	For pilus crystal agglutination (PCA)	
Anti-pilin IgG	62.5	2	
Anti-pilin F _{ab}	>500	>1,000	
Anti-28K IgG	62.5	31	
Anti-28K F _{ab}	62.5	>1,000	
BSA	>500	>1,000	

The table shows that the major pilus subunit (pilin) is antigenically unrelated to the 28K minor pilus protein. The level of antibodies against 28K elicited by immunization with intact pili is below the sensitivity of this ELISA. Anti-28K protein antibodies either bind too weakly to pili to resist the ELISA washing steps or are blocked from binding to the plastic-bound pili. Fab fragments were prepared from IgG antibodies purified from rabbit anti-pilin rod and anti-28K protein sera. The HA-inhibition by these antibodies is shown. The maximum concentration of antibodies or bovine serum albumin (BSA) tested was 500 μ g ml⁻¹ for inhibition assays, or 1,000 μ g ml⁻¹ for agglutination assays. Antisera against pili and pilus proteins were raised in rabbits and mice by subcutaneous injection of antigens at three- or four-week intervals. IgG was purified from rabbit serum by chromatography on DEAE-Affigel Blue (Biorad) and concentrated by ammonium sulphate precipitation and with a Centricon-10 (Amicon) microconcentrator. Purity was judged to be >98% by SDS-PAGE. Fab fragments were prepared by papain digestion of IgG, followed by chromatography on carboxymethylcellulose in acetate buffer at pH 5.5. Antibody activities were determined by ELISA, pilus paracrystalline aggregate agglutination^{6,14} and haemagglutination inhibition. An indirect ELISA (ref. 20) was done using peroxidase-conjugated second antibodies. ELISA titres were determined from single-reciprocal plots of absorbances from the enzyme/substrate/dye reaction (C.C.B. and J. R. Bryan, unpublished and ref. 6). IgGs and Fabs were serially diluted twofold and mixed with pilus crystals (10 μ g ml⁻¹) in 0.1 M MgCl₂. After determination of pilus crystal agglutination (PCA) end-points by dark-field microscopy, guinea pig erythrocytes were added to the pilus/antibody mixtures and allowed to settle in round-bottomed microtitre plates for one hour at 37 °C. Haemagglutination inhibition (HAI) end-points were determined visually. All IgG and Fab preparations used for HAI and PCA assays were active against their homologous antigens by ELISA (data not shown).

to erythrocyte membranes in an end-on fashion (data not shown), consistent with a tip location for pilus adhesive sites and the 28K protein.

Recent immunological electron microscope evidence indicates that the adhesin and two other minor proteins of the related Pap-pilus are at the cell-distal tips of these pili¹⁵. Thus, type-1 and Pap-pili seem to share the strategy, also seen in filamentous DNA phages¹⁶, of segregating adhesive function to minor subunits at their tips. In contrast, the adhesive function of some other *E. coli* pilus families seems to reside in the major pilin subunit^{17,18}.

Received 19 May 1987; revised 6 January; accepted 13 January 1988.

1. Brinton, C. C. Jr *Nature* **183**, 782-786 (1959).
2. Brinton, C. C. Jr *Trans. N.Y. Acad. Sci.* **27**, 1003-1054 (1965).
3. Old, D. C. *J. gen. Microbiol.* **71**, 149-157 (1972).
4. Ofek, I., Mirelman, D. & Sharon, S. *Nature* **265**, 623-625 (1977).
5. Brinton, C. C. Jr *et al. Vet. Med. Sm. Animal. Clin.* **78**, 962-966 (1983).
6. Fusco, P. thesis, Univ. Pittsburgh (1983).
7. Brinton, C. C. Jr *Proc. XIII U.S. - Japan Conf. on Cholera, Atlanta* p. 33-60 (1978).
8. Hagberg, L. *et al. Infect. Immunity* **40**, 265-272 (1983).
9. Swaney, L. M. *et al. J. Bact.* **130**, 495-505 (1978).
10. Maurer, L. & Orndorff, P. E. *FEMS Microbiol. Lett.* **30**, 59-66 (1985).
11. Maurer, L. & Orndorff, P. E. *J. Bact.* **169**, 640-645 (1987).
12. Markwell, M. K. *Analyt. Biochem.* **125**, 427-432 (1982).
13. Morrissey, J. H. *Analyt. Biochem.* **117**, 307-310 (1981).

The serological conservation of the 28K protein suggests that it could be useful as a component of vaccines against type 1-piliated pathogens. It is likely that functional constraints on other minor pilus proteins proposed to be adhesins^{15,19} have caused them to be serologically conserved as well. Purified pilus adhesins may prove to be effective as broad specificity, colonization-inhibiting vaccines.

We thank Jim Bryan for assistance with the ELISAs, Peter Fusco and Hyo-Jeong Cho for pili from the pathogenic strains, and Roger Hendrix for the phage lambda strains. We also thank Paul Orndorff for the type-1 pilus insertion mutants.

14. Hanson, M. S. thesis, Univ. Pittsburgh (1983).
15. Lindberg, F., Lund, B., Johansson, L. & Normark, S. *Nature* **328**, 84-87 (1987).
16. Grant, R. A., Lin, T.-C., Konigsberg, W. & Webster, R. E. *J. biol. Chem.* **256**, 539-546 (1981).
17. Jacobs, A. A. C., Venema, J., Leeven, R., van Pelt-Heerschap, H. & deGraaf, F. K. *J. Bact.* **169**, 735-741 (1987).
18. Jacobs, A. A. C., Simons, B. & deGraaf, F. J. *J. Bact.* **169**, 735-741 (1987).
19. Moch, T., Hochschützky, H., Hacker, J., Kröncke, K.-D. & Jann, K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3462-3466 (1987).
20. Voller, A., Bartlett, A. & Bidwell, D. E. *J. Clin. Path.* **31**, 507-520 (1978).
21. Fusco, P., To, C.-C., To, C.-M. & Brinton, C. C. Jr *Proc. XIII U.S.-Japan Conf. on Cholera, Atlanta* p. 60-69 (1978).
22. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
23. deGraaf, F. K. & Roorda, I. *Infect. Immunity* **36**, 751-758 (1982).
24. Towbin, H., Staehlin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).

Von Hippel-Lindau disease maps to the region of chromosome 3 associated with renal cell carcinoma

B. R. Seizinger¹, G. A. Rouleau¹, L. J. Ozelius¹, A. H. Lane^{1,2}, G. E. Farmer^{1,2}, J. M. Lamiell³, J. Haines⁴, J. W. M. Yuen⁵, D. Collins⁶, D. Majoor-Krakauer⁷, T. Bonner⁸, C. Mathew⁹, A. Rubenstein¹⁰, J. Halperin¹⁰, A. McConkie-Rosell¹¹, J. S. Green¹², J. A. Trofatter⁴, B. A. Ponder⁹, L. Eierman¹³, M. I. Bowmer¹², R. Schimke⁶, B. Oostra⁷, N. Aronin¹⁴, D. I. Smith¹⁵, H. Drabkin¹⁶, M. H. Waziri¹⁷, W. J. Hobbs¹, R. L. Martuza², P. M. Conneally⁴, Y. E. Hsia⁵ & J. F. Gusella¹

Von Hippel-Lindau disease (VHL) is an autosomal dominant disorder with inherited susceptibility to various forms of cancer, including hemangioblastomas of the central nervous system, pheochromocytomas, pancreatic malignancies, and renal cell carcinomas¹⁻³. Renal cell carcinomas constitute a particularly frequent cause of death in this disorder, occurring as bilateral and multifocal tumours, and presenting at an earlier age than in sporadic, non-familial cases of this tumour type. We report here that the VHL gene is linked to the locus encoding the human homologue of the *RAF1* oncogene, which maps to chromosome 3p25 (ref. 4). Crossovers with the VHL locus suggest that the defect responsible for the VHL phenotype is not a mutation in the

RAF1 gene itself. An alternative or prior event to oncogene activation in tumour formation may be the inactivation of a putative 'tumour suppressor' which can be associated with both the inherited and sporadic forms of the cancer. Sporadic renal cell carcinomas have previously been associated with the loss of regions on chromosome 3p (refs 5, 6). Consequently, sporadic and VHL-associated forms of renal cell carcinoma might both result from alterations causing loss of function of the same 'tumour suppressor' gene on this chromosome.

The primary biochemical defect in VHL is not yet known. We have therefore used genetic-linkage analysis with polymorphic DNA markers as a first step in applying chromosome-specific cloning techniques to the isolation and characterization of the defect. As VHL is associated with inherited susceptibility to cancer, mutations of oncogenes may directly cause the disease phenotype. A 2.9 kilobase (kb) complementary-DNA probe, p627 for the human *RAF1* oncogene locus on chromosome 3p25 (ref. 7), detects restriction fragment length polymorphism (RFLP) with the enzymes *TaqI* and *BglI*. With *TaqI*, the probe detects constant fragments of 3.9, 2.6, 2.5, 2.0, 1.9, 1.7, 0.9 and 0.6 kb, along with allelic fragments of 6.8 and 6.3 kb, having frequencies of 74% and 26%, respectively ($N = 116$ chromosomes). The 4.0 and 3.3-kb allelic fragments of the *BglI* RFLP have frequencies of 54% and 46% respectively ($N = 116$). Invariant fragments of 8.1, 6.4, 2.2, 0.6 and 0.5 kb are also detected in *BglI*-digested DNA. Haplotype frequencies for the presence or absence of the *TaqI* and *BglI* sites, respectively, were: '--', 42%; '-+', 26%; '+-', 10%; and '++', 22%, indicating only slight linkage disequilibrium between the two sites.

We assessed the potential genetic linkage of *RAF1* to VHL for all nine families using the computer program MLINK (V3.5) of the LINKAGE package⁸ with an age-of-onset correction^{2,9}. The results of this analysis are presented as lod scores in Table

¹ Neurogenetics Laboratory, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA.

² Neurosurgery Service, Department of Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA.

³ Intensive Care Unit, Letterman Army Medical Center, Residio of San Francisco, California 94129, USA.

⁴ Department of Medical Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46223, USA.

⁵ Department of Genetics and Pediatrics, University of Hawaii School of Medicine, Honolulu, Hawaii 96826, USA.

⁶ Department of Medicine, University of Kansas College of Health Sciences and Hospital, Kansas City, Kansas 66103, USA.

⁷ Department of Clinical Genetics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands.

⁸ Laboratory for Cell Biology, National Institute of Mental Health, Bethesda, Maryland 20892, USA.

⁹ Institute of Cancer Research, Royal Cancer Hospital, Sutton, Surrey, SM2 5PX, UK.

¹⁰ Department of Neurology, Mount Sinai School of Medicine, New York, New York 10029, USA.

¹¹ Department of Pediatrics, Duke University Medical Center, Durham, North Carolina 27710, USA.

¹² Health Sciences Center, Memorial University, St John's, Newfoundland A1B 3V6, Canada.

¹³ Department of Pediatrics, VMD, Division of Medical Genetics, Rutgers Medical School, New Brunswick, New Jersey 08903, USA.

¹⁴ Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA.

¹⁵ Department of Molecular Biology and Genetics, Wayne State University, Detroit, Michigan 48202, USA.

¹⁶ Department of Medicine, University of Colorado, Denver, Colorado 80262, USA.

¹⁷ Department of Pediatrics, Division of Medical Genetics, University Hospital, Iowa City, Iowa 52242, USA.

Table 1 Lod scores for linkage of VHL to the *RAF1* gene

VHL pedigree	Recombination fraction (θ)									Peak lod score	
	0.00	0.01	0.05	0.10	0.15	0.20	0.30	0.40	($\hat{z}$)	at $\hat{\theta}$	
1	1.95	1.92	1.77	1.58	1.38	1.17	0.73	0.27	1.95	0.00	
2	$-\alpha$	-1.63	-0.34	0.13	0.33	0.42	0.43	0.27	0.45	0.25	
3	0.73	0.71	0.64	0.55	0.46	0.37	0.19	0.05	0.73	0.00	
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	—	—	
5	0.39	0.38	0.35	0.29	0.24	0.18	0.08	0.02	0.39	0.00	
6	0.63	0.61	0.55	0.46	0.38	0.30	0.15	0.04	0.63	0.00	
7	$-\alpha$	-1.35	-0.63	-0.32	-0.15	-0.04	0.05	0.06	0.06	0.40	
8	$-\alpha$	0.17	0.73	0.85	0.84	0.77	0.51	0.18	0.86	0.12	
9	0.80	0.82	0.85	0.83	0.78	0.69	0.46	0.17	0.85	0.06	
All families	$-\alpha$	1.63	3.92	4.37	4.26	3.86	2.60	1.06	4.38	0.11	

We have established permanent lymphoblastoid cell lines¹² from 203 members of nine families with VHL, including 71 affected individuals. The diagnosis of VHL in these families was made using the criteria of Melmon and Rosen¹. Affected individuals from all nine families displayed typical VHL, including hemangioblastomas in the cerebellum, retina, and other parts of the central nervous system, renal cell carcinomas, pancreatic tumours, and multiple cysts in kidney, pancreas, and epididymis¹⁻³. Phaeochromocytomas, however, which can also be associated with VHL were seen only in pedigree number 7. Pedigrees 1 and 7 have been described in detail elsewhere^{1,13}. **Methods.** DNA was extracted either from fresh blood samples¹⁴ or from lymphoblastoid cell lines¹⁵. For RFLP typing, ~5 µg of each DNA was digested to completion with the restriction enzymes *TaqI* or *BglI*. The resulting fragments were resolved by agarose-gel electrophoresis, transferred to nylon membrane by the Southern blot technique, and hybridized with ³²P-labelled probe p627 as described¹⁶. Alleles present in each individual were inferred from the sizes of bands on the developed autoradiograph. Linkage analysis was carried out using the computer program MLINK (V3.5) of the LINKAGE⁸ package with an age-of-onset correction consisting of ten age-dependent penetrance classes based on the age-of-onset curve for 363 VHL cases by Go *et al.*². A penetrance of 95% was assumed for all ages over 44 years, with a frequency of 0.0001 for the disease gene. Altering this penetrance to 90% or 100% did not significantly change the overall lodscores. The results are expressed as a lod score (z) representing the log₁₀ of the ratio of the likelihood of linkage to VHL at the specified recombination fractions (θ) relative to the likelihood of non linkage ($\theta = 0.50$). Generally, a lod score of $>+3$ is considered proof of linkage whereas a lod score of <-2 excludes linkage at the specified θ value.

1. A combined maximum lod score ($\hat{z}$) of 4.38 was obtained at $\hat{\theta} = 0.11$ with a 1-lod unit confidence interval (approximating a 95% confidence interval) with a θ value of 0.04–0.23. None of the families provided significant evidence of non-allelic heterogeneity. These data clearly establish that the defect causing VHL is located on the short arm of chromosome 3. But although the VHL defect and *RAF1* are genetically linked, crossovers between the two loci suggest that a mutation in the *RAF1* gene itself is unlikely to be the fundamental defect responsible for the VHL phenotype.

Knudson¹⁰ has proposed that sporadic and inherited forms of particular cancers might be caused by mutations in the same gene. In this model, both copies of a putative 'tumour suppressor' gene must be lost or inactivated to induce tumour formation. In familial cases, susceptibility to tumour development is inherited through a germ-line mutation and the tumour develops after the normal balancing wild-type gene is lost or inactivated by a somatic change, thereby unmasking the recessive mutant allele. In sporadic cases both inactivation events represent somatic alterations. Karyotype and molecular-genetic studies of a number of sporadic renal cell carcinomas and of the renal cell carcinoma from a single VHL patient have revealed deletions on chromosome 3p^{5,6,11}. As the VHL gene maps to this region on chromosome 3, it may encode a 'tumour suppression' whose loss of function is of fundamental importance for the development of several inherited types of human cancer and their sporadic counterparts.

We thank Dr F. Li for helpful discussions, A. G. Faryniarz, J. A. Iannazzi, J. C. Roy, T. J. Corey and M. A. Anderson for technical assistance, and Karen Griffen and Christine Reuss for typing the manuscript. B.R.S. is supported by a Junior Investigator Award from the National Neurofibromatosis Foundation, G.A.R. is supported in part by the Fonds de Recherche en Sante du Quebec and is a Fellow of the Medical Research Council of Canada. J.F.G. is a Searle Scholar of the Chicago Community Trust. This work was supported by the National Institute for Neurological and Communicative Disorders and Stroke, the McKnight Foundation, the Hereditary Disease Foundation, the Dystonia Medical Research Foundation, and the Julieanne Dorn Fund for Neurological Research.

3. Huson, S. M. *et al. Brain* **109**, 1297–1310 (1986).

4. Bonner, T. *et al. Science* **223**, 71–74 (1984).

5. Yoshida, M. A. *et al. Cancer Genet. Cytogenet.* **19**, 351–354 (1986).

6. Zbar, B., Brauch, H., Talmadge, C. & Linehan, M. *Nature* **327**, 721–724 (1987).

7. Bonner, T. I. *et al. Nucleic Acids Res.* **14**, 1009–1015 (1986).

8. Lathrop, G. M. & Lalouel, J. M. *Am. J. hum. Genet.* **36**, 460–465 (1984).

9. Hodge, S. E., Morton, L. A., Tideman, S., Kidd, K. K. & Spence, M. A. *Am. J. hum. Genet.* **31**, 761–762 (1979).

10. Knudson, A. G. *Cancer Res.* **45**, 1437–1443 (1985).

11. King, C. R. *et al. Cancer Genet. Cytogenet.* **27**, 345–348 (1987).

12. Gusella, J. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 5239–5243 (1979).

13. Green, J. S., Bowmer, M. I. & Johnson, G. J. *Can. med. Ass. J.* **134**, 133–138 (1986).

14. Seizinger, B. R., Martuza, R. L. M. & Gusella, J. F. *Nature* **322**, 644–647 (1986).

15. Anderson, M. A. & Gusella, J. F. *In Vitro* **11**, 856–858 (1984).

16. Seizinger, B. R. *et al. Cell* **49**, 589–594 (1987).

Sequence similarity of phospholipase C with the non-catalytic region of src

Mark L. Stahl, Catherine R. Ferenz, Kerry L. Kelleher, Ron W. Kriz & John L. Knopf

Department of Molecular Biology, Genetics Institute Inc., 87 Cambridge Park Drive, Cambridge, Massachusetts 02142, USA

The production of the second messenger molecules diacylglycerol and inositol 1,4,5-trisphosphate¹ is mediated by activated phosphatidylinositol-specific phospholipase C (PLC) enzymes^{2,3}. Here we report the cloning of a bovine brain complementary DNA encoding an enzyme PLC-148 that is characterized by calcium-dependent and phosphatidylinositol-specific phospholipase C activity when expressed in mammalian cells. Bovine brain messenger RNA contains a 7.5-kilobase transcript corresponding to the isolated cDNA; a related transcript of the same size is present in mRNA from some but not all human cell lines tested. Southern blot analysis of the bovine genome indicated that one or possibly two genes hybridize to the cloned PLC-148 cDNA. There is a striking sequence similarity between specific regions of PLC-148 and the non-catalytic domain of the non-receptor tyrosine kinases. The newly characterized *crk* transforming gene of the avian sarcoma virus CT10 also contains extensive sequence similarities with PLC-148.

Accumulated evidence suggests the occurrence of a number of distinct enzymes with phosphatidylinositol-specific PLC activities^{4,5}. Because of the importance of PLC enzymes in the

Received 18 December 1987; accepted 28 January 1988.

1. Melmon, K. L. & Rosen, S. W. *Am. J. Med.* **36**, 595–617 (1964).

2. Go, R. C. P., Lamiell, J. M., Hsia, Y. E., Yuen, J. W. M. & Paik, Y. *Am. J. hum. Genet.* **36**, 131–142 (1984).

Bovine brain mRNA was analysed for species corresponding to the PLC-148 cDNA clone. A single 7.5-kb transcript was detected (Fig. 2a, lane 5). Transcripts of the same size were detected in RNA from several human cell lines, including WI-38 cells (fibroblasts), LS174 cells (colon carcinoma) and U20S cells (osteosarcoma). No cross-hybridizing mRNA was detectable in

Fig. 2 *a*, Northern blot analysis of human poly(A)⁺ mRNA (5 µg- derived from the fibroblast cell line WI-38 (lane 1), the promyelocyte line HL-60 (lane 2), the colon carcinoma line LS174 (lane 3), osteosarcoma cell line U2OS (lane 4), and bovine brain (lane 5), performed essentially as described¹³. The nitrocellulose was hybridized with a ³²P-labelled single-stranded DNA probe (extending from nucleotide 2,855-4,312 in Fig. 1) at 60 °C in 4×SSC, 5×Denhardt's, 100 µg denatured calf-thymus DNA, 0.1% SDS for 14 h, and washed at 60 °C in 4×SSC, 0.1% SDS for 1 h and autoradiographed for 2 days at -70 °C. Ribosomal RNA size markers are indicated. *b*, Southern blot analysis of bovine genomic DNA. DNA (10 µg) was digested with *Bgl*III (lane 1), *Eco*RI (lane 2), *Hind*III (lane 3), *Pst*I (lane 4), and *Sac*I (lane 5). Hybridization was as described above. DNA size markers from the top of the blot are 23,130, 9,419, 6,557, 4,371 and 2,322 base pairs respectively.

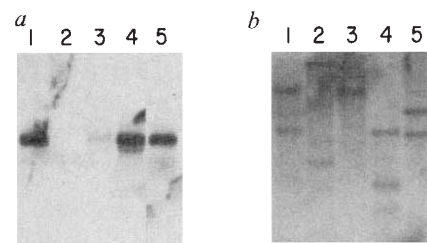


Fig. 3 *a*, Linear representation of *c-src*, PLC-148, and *gag-crk*. The regions displaying sequence similarity are indicated. The numbers below designate the amino-acid residue numbers and identify these regions in *b*. *b*, Amino acid alignment of the conserved amino-terminal region of cytoplasmic tyrosine kinases. The figure is modelled after that presented in ref. 7: a residue appears in upper case if present in the same position in three or more of the sequences, and in lower case if present in less than three of the presented sequences. Amino-acid residues and blocks of residues strongly conserved in the tyrosine kinases are boxed. These boxed regions are extended to indicate the corresponding region(s) of PLC and *crk*. These same conserved regions are found in both *crk* and PLC-148. The sequences labelled *v-fgr* and *v-abl* begin at the junction of their respective *gag/kinase* fusion. The sequences are divided into a proposed modulatory domain and a catalytic domain where sequence homology with the receptor tyrosine kinases begins. The amino acid sequences of the proteins encoded by the following are presented: *crk* (ref. 8), *v-src* (ref. 15), *c-src* (ref. 16), *v-yes* (ref. 17), *Lck* (ref. 18), *syn* (ref. 19), *hck* (ref. 20), *lyn* (ref. 21), *v-fgr* (ref. 22), *v-abl* (ref. 23).

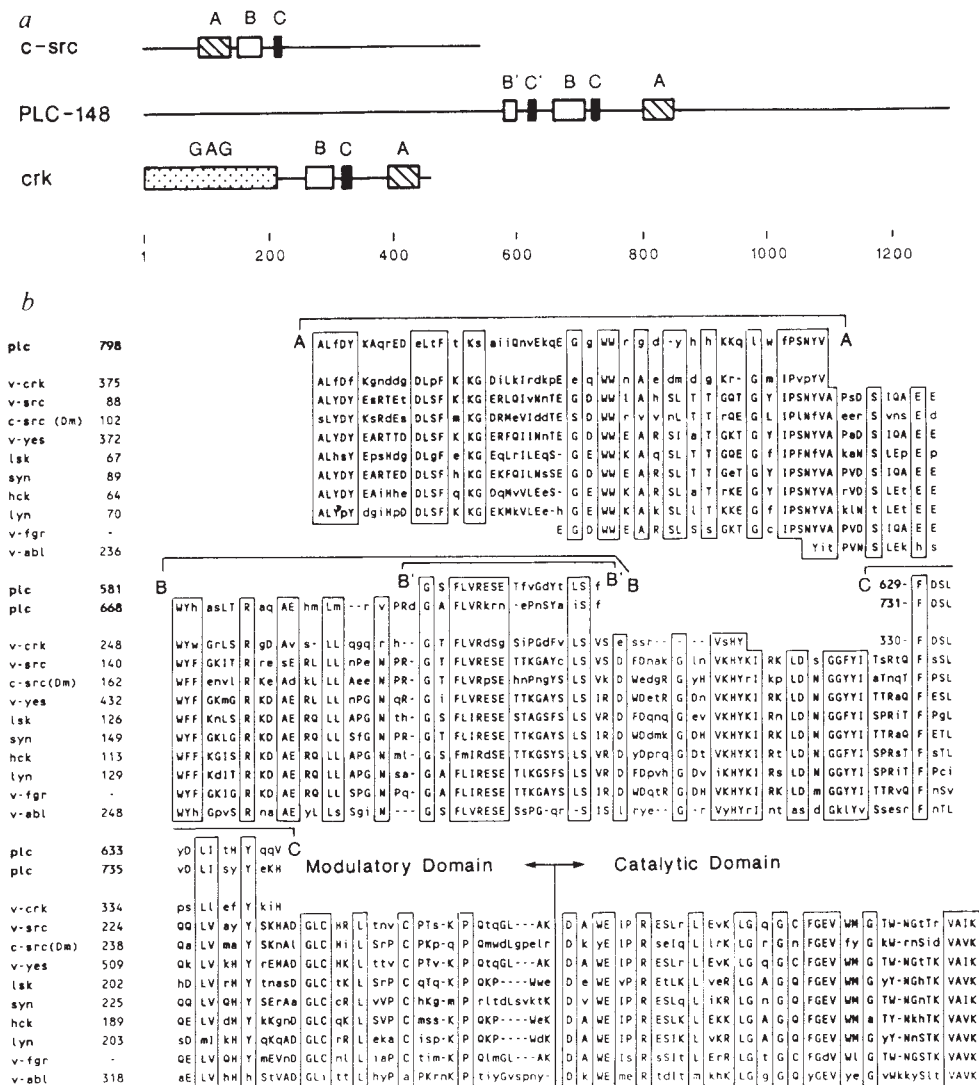


Table 1 Phospholipase C activity in transfected COS-1 cells

Transfected DNA	Assay conditions -Ca ²⁺	+Ca ²⁺
pmt	<0.01	2.0
pmt-PLC	<0.01	78

The cDNA insert shown in Fig. 1 was subcloned into pmt-PLC, the COS cell expression vector pmt ampicillin derivative of p91023. The indicated plasmid DNA (32 µg) was transfected into four 10-cm dishes¹⁴. 60 h after the addition of the DNA, cells were lysed in homogenization buffer⁶ and spun at 100,000g for 60 min. The supernatant was assayed for PLC activity in the presence of 0.1% deoxycholate using [³H]phosphatidylinositol (New England Nuclear)⁶. Numbers represent c.p.m. × 10⁵ of [³H]inositol released into the aqueous phase per 4 dishes of cells per 10 minutes.

the human promyelocytic cell line, HL60 (Fig. 2a, lane 2). A unique mRNA encoding the PLC is thus highly conserved between cow and man.

We also used PLC-148 cDNA to probe Southern blots of bovine genomic DNA (Fig. 2b). The digestion patterns resulting from five restriction endonucleases (*Bgl*III, *Eco*RI, *Hind*III, *Pst*I and *Sac*I) suggest that only one, or possibly two genes hybridize to the probe.

A search of the NBRF database revealed a significant homology between PLC-148 and the non-receptor protein tyrosine kinases. Three clustered regions of homology with *src* were noted and are referred to as A, B and C. The relative location of these regions within *src* and PLC-148 is shown in Fig. 3a. The *src* A, B and C regions are contained in the conserved amino-terminal, non-catalytic domain of the protein⁷. The corresponding regions of PLC-148 appear in the central portion of

the enzyme. PLC-148 has two B-like regions, referred to as PLC B and B', as well as two C-like regions, PLC C and C'. There is only one PLC A region which is transposed from its relative position in *src* to the carboxy-terminal side of the PLC B and C regions. Together the three blocks of sequence similar to PLC-148 contain >50% of the non-catalytic amino-terminal domain which is conserved among the cytoplasmic tyrosine kinases. Recently we have learned that the transforming gene *crk* of the avian sarcoma virus CT10 (CT10 regulator of kinase) is also homologous to the same regions of the *src*-encoded family of tyrosine kinases⁸. Interestingly, the arrangement of the three *crk*-encoded regions has more similarity to the arrangement found in PLC-148 than to the tyrosine kinases, because the *crk*-B and C regions precede the *crk*-A region. The spacings of A, B and C are also similar between *crk* and PLC-148. An alignment of the sequences contained in the A, B and C regions of both PLC-148 and *crk* with a number of cytoplasmic tyrosine kinases is shown in Fig. 3b. The residues and blocks of residues which are strongly conserved in the amino-terminal domain of the tyrosine kinases⁷ are often conserved in both PLC-148 and *crk*.

The function of the domain common to *crk* and PLC-148 is uncertain at present. It is possible that *crk* encodes a constitutively activated catalytic domain of a phospholipase (related to PLC-148) which could transform cells. At present we consider this unlikely because of the extensive sequence similarity between this domain and the non-catalytic domain of the tyrosine kinases. Possibly the domain shared by *crk* and PLC-148 serves a regulatory function in PLC-148, as already suggested for the corresponding domain in the tyrosine kinases^{7,9,10}, and supported by a study of mutations in the *fps* amino-terminal region affecting both the tyrosine kinase activity and the transforming ability of the mutant virus⁷. The regulation of kinase and transforming functions possibly requires this domain to interact with cellular components which together modulate these activities⁷, possibly by the phosphorylation of tyrosine residues^{8,9}. The *crk* gene product has no homology with a tyrosine kinase catalytic domain, but can still modulate the activity of cellular tyrosine kinases either directly or indirectly by mechanisms possibly involving the titration of an inhibitor⁸. The corresponding region of PLC-148 could modulate PLC activity and along with the tyrosine kinases and the *crk* proto-oncogene could possibly be recognized by or interact with similar cellular components. Cellular components so far known to regulate the activity of some PLCs are members of the G-protein family¹, which have not been shown to interact with cytoplasmic tyrosine kinases. The recombinant enzyme will enable the possible regulation of PLC-148 by GTP-binding proteins to be studied. Our observations suggest that potential regulatory domains could serve as an initial focus for the generation of mutant enzymes and the study of the consequences of these mutations *in vivo* could provide an indication of the function, if any, of this enzyme in signal transduction.

We thank Richard Jove for suggesting that PLC may be structurally related to *crk*, Jasbir Sehra, Rodney Hewick and Robert Kamen for discussions, Timothy O'Toole and Janet Judd for oligonucleotide synthesis, and Clive Wood and Robert Kamen for critical comments on the manuscript.

Received 30 December 1987; accepted 1 February 1988.

- Berridge, M. J. *Biochem. J.* **220**, 345-360 (1984).
- Berridge, M. J. *A. Rev. Biochem.* **56**, 159-194 (1987).
- Majerus, M. J. *Science* **234**, 1519-1526 (1986).
- Ryu, S. H., Suh, P., Cho, K. S., Lee, K. Y. & Rhee, S. G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6649-6653 (1987).
- Hofmann, S. L. & Majerus, P. W. *J. biol. Chem.* **257**, 6461-6469 (1982).
- Ryu, H. R., Cho, K. S., Suh, P. & Rhee, S. G. *Biochem. biophys. Res. Commun.* **141**, 137-144 (1986).
- Sadowski, I., Stone, J. & Pawson, T. *Molec. cell. Biol.* **6**, 4396-4408 (1986).
- Mayer, B. J., Hamaguchi, M. & Hanafusa, H. *Nature*, this issue.
- Jove, R. & Hanafusa, H. *An. Rev. cell Biol.* **3**, 31-56 (1987).
- McCarley, D. J., Parsons, J. T., Benjamin, D. C. & Parsons, S. J. *J. Virol.* **61**, 1927-1937 (1987).
- Kozak, M. *Nucleic Acids Res.* **20**, 8125-8148 (1987).
- Jacobs, K. *et al. Nature* **313**, 806-810 (1985).

- Toole, J. J. *et al. Nature* **312**, 342-347 (1984).
- Knopf, J. L. *et al. Cell* **46**, 491-502 (1986).
- Takeya, T., Feldman, R. A. & Hanafusa, H. *J. Virol.* **44**, 1-12 (1982).
- Simon, M. A., Drees, B., Kornberg, T. & Bishop, M. J. *Cell* **42**, 831-840 (1985).
- Kitamura, N. *et al. Nature* **297**, 205-208 (1982).
- Marth, J. D., Peet, R., Krebs, E. G. & Perlmutter, R. M. *Cell* **43**, 393-404 (1985).
- Semba, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 5459-5955 (1986).
- Quintrell, N. *et al. Molec. cell. Biol.* **7**, 2267-2275 (1987).
- Yamanashi, Y. *et al. Molec. cell. Biol.* **7**, 237-243 (1987).
- Naharro, G. K., Robbins, K. C. & Reddy, E. P. *Science* **223**, 613-66 (1984).
- Reddy, E. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 3624-3628 (1983).

A novel viral oncogene with structural similarity to phospholipase C

Bruce J. Mayer, Michinari Hamaguchi
& Hidesaburo Hanafusa

The Rockefeller University, New York, New York 10021, USA

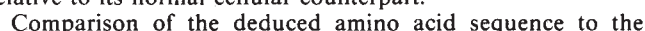
Numerous oncogenes have been isolated from acutely transforming retroviruses. To date, the products of these viral oncogenes have been protein kinases, nuclear proteins, growth factors, or GTP-binding proteins^{1,2}. We have cloned the previously uncharacterized avian sarcoma virus CT10 and sequenced its genome. This virus encodes a protein, p47^{gag-cr_k}, that has blocks of sequence similarity to the amino-terminal, non-catalytic region of the non-receptor class of tyrosine kinases. In addition, the structure of p47^{gag-cr_k} has striking similarity to a 180-amino acid region of bovine brain phospholipase C. Biochemical data suggest that p47^{gag-cr_k} activates one or several endogenous tyrosine kinases.

We have characterized the avian sarcoma virus CT10 (refs 3, 4) which was passaged in chicken embryo fibroblasts (CEF) to generate high-titre stocks. The virus caused CEF to assume an elongated, fusiform morphology and to grow to high density. When injected into newborn chickens, the virus caused rapid tumour formation (within two weeks) at the site of injection. When ³²P-labelled complementary DNA made from partially purified CT10 virion RNA was used as a probe, no hybridization was observed to a battery of 19 DNA fragments isolated from molecularly cloned viral oncogenes. This suggested that CT10 contained a novel oncogene, so we decided to molecularly clone and sequence the CT10 genome.

Northern blotting of CT10 virus-infected cell RNA probed with DNA fragments derived from UR2AV, a cloned avian leukosis virus⁵, demonstrated that the CT10 genomic RNA was ~2.5 kilobases (kb) long and contained sequences homologous to the long terminal repeat (LTR) and 5' *gag* regions of UR2AV only (not shown). Intact helper virus (termed CT10AV) was also present in stocks, presumably to provide essential replicative gene products to the defective sarcoma virus *in trans*.

We have isolated unintegrated viral DNA from infected QT6 quail cells, mapped restriction sites in the CT10 genome by Southern blotting, and cloned the permuted circular genome into a lambda phagemid vector⁵⁻⁷. Five molecular clones were isolated that contained inserts of ~2.4 kb and hybridized with LTR and 5' *gag* probes only. Two independent clones that had ligated into the vector in opposite orientations were chosen for further analysis. Both clones, when transfected into CEF (ref. 8) with UR2AV helper virus DNA, induced transformed morphology indistinguishable from that induced by the parental CT10 virus. Culture fluids from transfected cells contained infectious virus that rapidly caused tumours when injected into chickens. These results demonstrate that we had cloned biologically active sarcoma virus.

The nucleotide sequence of the cloned viral DNA and the amino acid sequence of its putative protein product are presented in Fig. 1. The nucleotide sequences of the two cloned isolates were identical except for three bases in the non-coding *gag*



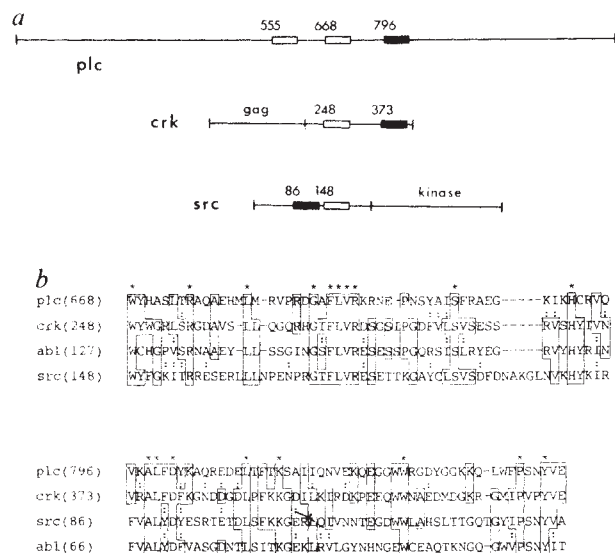


Fig. 2 Structural similarities between CT10 transforming protein (p47^{gag-crk}), phospholipase C, and the *c-src* and *c-abl* gene products. **a**, Comparison of the structure of phospholipase C-148 (ref. 15), p47^{gag-crk}, and pp60^{c-src} (ref. 21). Open and black boxes each denote a region of ~50 amino acids of similar sequence (termed regions B and A respectively by M. Stahl *et al.*¹⁵). Numbers refer to initial amino acid of homology box. **b**, Sequence comparison of similar regions indicated above, with the addition of *c-abl* (ref. 22). Boxed amino acids are identical to p47^{gag-crk}; asterisks denote residues conserved in all four sequences, and colons indicate conservative amino acid changes. Gaps imposed to maximize alignment are indicated by dashes. Number in parentheses is the first amino acid of sequence shown.

protein database showed two 50-amino acid blocks of sequence similarity to the non-receptor class of tyrosine kinases (only one of these blocks is conserved in *c-fps*). One of these blocks corresponds to the SH2 domain initially identified by Sadowski *et al.*²⁵ (Fig. 2). Two aspects of this homology were surprising (we use the term homology for significant sequence similarity). First, the homology is confined to the amino-terminal domain of the tyrosine kinases; the catalytic activity of the tyrosine kinases resides in a strongly conserved carboxy-terminal domain¹⁴ that is totally absent from the CT10 oncogene product. In addition, the two blocks of homologous sequence are trans-

posed and spaced quite differently relative to each other when the CT10 protein and the tyrosine kinases are compared (see Fig. 2a). At this time we received by personal communication the amino acid sequence of phosphatidylinositol-specific, calcium-dependent phospholipase C-148 isolated from bovine brain¹⁵. Comparison of phospholipase C (PLC) with the CT10 protein showed striking structural similarities (Fig. 2). The two proteins contain blocks of homology similar to the tyrosine kinases and the relative position and spacing of the elements is virtually identical, although the intervening sequences have only a short stretch of additional homology. This structural similarity raises the possibility that the transforming gene product of CT10 arose by transduction and truncation of a member of the phospholipase C family.

As the putative 47K oncogene product contains 208 amino acids of viral gag sequence, we immunoprecipitated it with antibodies raised against avian retrovirus gag protein. Using sera from rabbits bearing RSV-induced tumours (TBR sera) or injected with disrupted RAV-2 virions (anti-virion sera), a 47K protein was precipitated from [³H]leucine-labelled CT10-infected cell lysates (Fig. 3). This was true of cells infected with the original CT10 stock (Fig. 3a) or with molecularly cloned virus (Fig. 3b); however, the amount of p47 in the cells infected with cloned virus was much lower, presumably reflecting a difference in the ratio of sarcoma virus to helper virus in the different virus stocks. To our surprise, when an *in vitro* kinase assay was performed on TBR immunoprecipitates of CT10-infected cell lysates, a diffuse band of 135–155K was phosphorylated (Fig. 3c). This band, which was not present in the control, was resistant to alkaline hydrolysis, and phosphoamino-acid analysis confirmed that it was phosphorylated on tyrosine (not shown). Kinase assays performed on lysates from cells infected with the molecularly cloned virus gave similar results, but the phosphorylation of the 135–155K band was less pronounced, probably reflecting the lower levels of p47 present. It is unclear whether the observed tyrosine kinase activity is due to a direct association between p47 and a (presumably activated) cellular kinase, or whether the TBR serum cross-reacts with a cellular kinase whose activity has been greatly elevated in infected cells.

To assess the effect of p47 on the state of tyrosine phosphorylation *in vivo*, we probed an immunoblot of CT10-infected cell lysates with an antibody that specifically recognizes phosphotyrosine^{16,17}. Figure 4 shows that there is a dramatic increase in the tyrosine phosphorylation of a relatively small number of proteins in cells infected with molecularly cloned CT10 virus

Fig. 3 Immunoprecipitation of CT10-specific protein and *in vitro* kinase activity of immunoprecipitates. **a**, [³H]leucine-labelled proteins from CEF infected with original CT10 stock immunoprecipitated with TBR serum. Samples were separated on 8% SDS/polyacrylamide gel. CT10-specific p47 band is indicated, with major helper virus proteins p27, pr76 and pr180. Size markers (M from top of gel): 200, 97, 68, 43, and 26K. **b**, [³H]leucine-labelled proteins from CEF infected with molecularly cloned CT10 virus, immunoprecipitated with either TBR or anti-virion serum; 10% SDS/polyacrylamide gel. Size markers as in **a**. **c**, *In vitro* kinase assay performed on unlabelled CT10-infected or uninfected CEF lysates; reaction products run on 8% gel. Major *in vitro* phosphorylated band of 135–155K is indicated. Size markers as in **a**.

Methods. Conditions for cell culture, labelling, preparation of cell lysates, immunoprecipitation, and kinase assay were essentially as described²³ with the following modifications. Lysates were prepared in complete RIPA buffer²⁶ containing 0.1% SDS and 1% sodium deoxycholate. Cells were labelled for 4 h with 1 mCi [³H]leucine (60 Ci mmol⁻¹; Amersham) in 1 ml. Kinase assays were performed in 50 mM HEPES (pH 7.4), 10 mM MnCl₂, 10 μCi [^γ-³²P]ATP (3,000 Ci mmol⁻¹; Amersham) for 10 min at room temperature; kinase assay gel was washed in 1 M KOH for 2 h at 55 °C to enrich for phosphotyrosine²⁴.

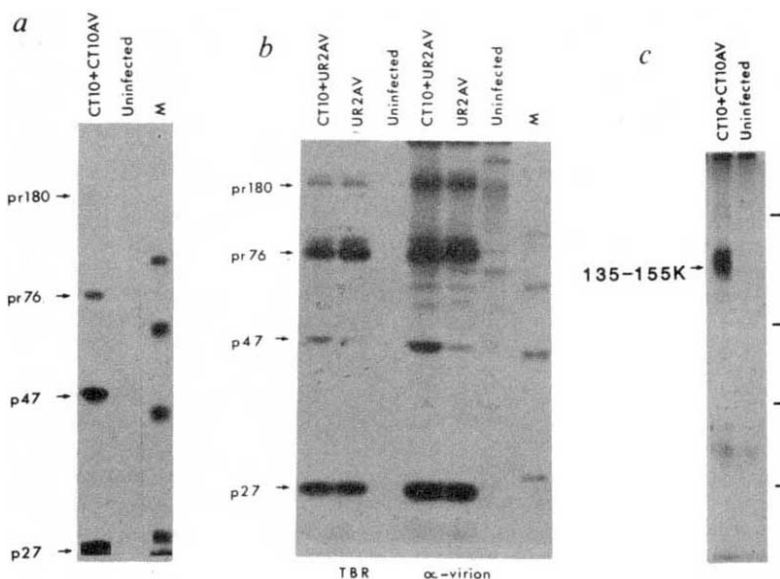
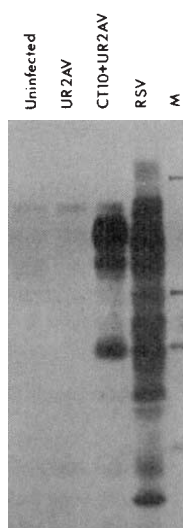


Fig. 4 Anti-phosphotyrosine immunoblot. Lysates from CEF infected with molecularly cloned CT10 virus or as indicated were separated by SDS-PAGE, transferred to nitrocellulose, and phosphotyrosine-containing proteins visualized by immunoradiography.

Methods. The method is described in detail elsewhere¹⁷. Briefly, cells were lysed by boiling in 1% SDS, 10 mM Tris (pH 7.4), 1 mM PMSF, 1 mM Na₂VO₄ and 0.1 mM Na₂MoO₄, lysates were run on 7.5% SDS/polyacrylamide gel, and proteins were transferred electrophoretically to nitrocellulose in Tris-glycine buffer. Filters were blocked in 10 mM Tris (pH 7.4), 0.1% Triton X-100, 0.9% NaCl with 1% ovalbumin (Sigma, fraction V), then incubated in blocking solution containing 0.75 µg ml⁻¹ affinity-purified antibody raised against phosphotyrosine, as described¹⁶. Filters were washed in blocking solution without ovalbumin, blocked again, incubated with 2 µCi [¹²⁵I]-protein A (~40 mCi mg⁻¹; Amersham), washed again and exposed. Comparable results were obtained using a completely different antibody raised against phosphotyramine¹⁷. Size markers are as in Fig. 3.



relative to helper-infected or uninfected cells. Furthermore, the pattern observed is very different from that seen in RSV-infected cells. The most prominent tyrosine phosphorylation induced by CT10 infection is in the 135–155K region, corresponding to the band observed in *in vitro* kinase assays. These results demonstrate that p47, a protein that does not contain any region homologous to the catalytic domain of known protein kinases, can modulate the apparent activity of endogenous tyrosine kinases *in vitro* and *in vivo*. We have therefore named this new oncogene *crk*, for CT10 regulator of kinase, and the oncogene product p47^{gag-crck}.

There are several implications of this work for mechanisms of signal transduction and growth control. As the catalytic site of PLC has not yet been identified, it is possible that p47^{gag-crck} encodes a functional lipase in the absence of normal regulatory domains, leading to constitutive phospholipid turnover. Such a mechanism would be consistent with numerous biochemical studies implicating the PLC-mediated cleavage of phosphatidylinositol-(4,5)bisphosphate in signal transduction pathways^{18,19}. It is also possible that p47^{gag-crck} transforms by direct or indirect modulation of endogenous tyrosine kinases, perhaps by associating in an activated complex or titrating away inhibitory factors. In any case, the fairly extensive homology between such functionally diverse enzymes as PLC and the non-receptor class of tyrosine kinases is intriguing, as it suggests a similar activity or interaction with a similar cellular component. The ability of the *crk* gene product to raise intracellular phosphotyrosine levels suggests that members of the PLC family and the non-receptor tyrosine kinases might interact in normal signal transduction pathways.

We thank Anindya Dutta and Sally Kornbluth for critically reading this manuscript, Ricardo Feldman for providing antiviral antibody, and David Foster, Carla Grandori, Teruko Hanafusa, Lu-Hai Wang and other members of the laboratory for contributions to this work. We also acknowledge the contribution of Richard Jove, who was the first to be aware of the possible relationship between PLC-148 and p47^{gag-crck}. Support for this work was provided by grants from the National Institutes of Health (to H.H.) and by a National Science Foundation graduate fellowship (to B.J.M.).

- Burk, D. *et al. J. natn. Cancer Inst.* **2**, 201–240 (1941).
- Claude, A., Porter, K. R. & Pickels, E. G. *Cancer Res.* **7**, 421–429 (1947).
- Neckameyer, W. S. & Wang, L.-H. *J. Virol.* **50**, 914–921 (1984).
- Mayer, B. J. *et al. J. Virol.* **60**, 858–867 (1986).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratories, 1982).
- Cross, F. R. & Hanafusa, H. *Cell* **34**, 597–607 (1983).
- Coffin, J. in *RNA Tumor Viruses, Molecular Biology of Tumor Viruses* Vol. 2 (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 17–73 (Cold Spring Harbor Laboratories, 1985).
- Takeya, T., Feldman, R. A. & Hanafusa, H. *J. Virol.* **44**, 1–11 (1982).
- Schwartz, D. E., Tizard, R. & Gilbert, W. *Cell* **32**, 853–869 (1983).
- Lerner, T. L. & Hanafusa, H. *J. Virol.* **49**, 549–556 (1984).
- Bizub, D., Katz, R. A. & Skalka, A. M. *J. Virol.* **49**, 557–565 (1984).
- Hunter, T. & Cooper, J. A. *Rev. Biochem.* **54**, 897–930 (1985).
- Stahl, M., Ferenz, C., Kelleher, K., Kriz, R. & Knopf, J. *Nature* **332**, 269–272 (1988).
- Wang, J. Y. *J. Molec. cell. Biol.* **5**, 3640–3643 (1985).
- Hamaguchi, M., Grandori, C. & Hanafusa, H. *Molec. cell. Biol.* (submitted).
- Berridge, M. J. A. *Rev. Biochem.* **56**, 159–193 (1987).
- Majerus, P. W. *et al. Science* **234**, 1519–1526 (1986).
- Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165–170 (1985).
- Takeya, T. & Hanafusa, H. *Cell* **32**, 881–890 (1983).
- Ben-Neriah, Y., Bernards, A., Paskind, M., Daley, G. Q. & Baltimore, D. *Cell* **44**, 577–586 (1986).
- Garber, E. A., Mayer, B. J., Jove, R. & Hanafusa, H. *J. Virol.* **61**, 354–360 (1987).
- Cooper, J. A. & Hunter, T. *Molec. cell. Biol.* **1**, 165–178 (1981).
- Sadowski, I., Stone, J. & Pawson, T. *Molec. cell. Biol.* **6**, 4396–4408 (1986).
- Collett, M. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021–2025 (1978).

A Harvey-ras responsive transcription element is also responsive to a tumour-promoter and to serum

J. L. Imler, C. Schatz, C. Wasylyk, B. Chatton & B. Wasylyk

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cédex, France

The *ras* oncogenes are implicated in the onset of some human tumours, and in cellular proliferation and terminal differentiation. The *ras* proteins are plasma membrane bound transducers of signals between the outside of the cell and unknown targets in the cell^{1,2}. Identifying these targets and understanding how they are regulated will have a major impact on our understanding of the molecular basis of transformation. We have already shown that *c-Ha-ras* and the tumor promoter TPA (12-*o*-tetradecanoyl phorbol-13-acetate) can activate a transcriptional enhancer³. We now report the identification of a short sequence in the polyoma virus (Py) enhancer which mediates Ha-ras activation, and show that this sequence (ras responsive element, RRE) also mediates activation by TPA and serum. This responsive element is a specific binding-site for the mouse transcription factor PEA1 (ref. 4 and below) and for the *jun* oncogene (ref. 5 and M. Karin, personal communication). These results are in keeping with a role for *ras* protein in signal transduction from outside the cell to a transcription factor in the nucleus, through protein kinase C^{1,2}. The striking similarity between RRE and DNA sequences present in the promoter regions of a number of transformation-related genes suggests that deregulated activation of RRE is a critical event in transformation.

Our aim is to identify the transcription factors whose activities are regulated by the *ras* oncogene. We initially studied the α domain of the Py enhancer⁴ (PyA region Fig. 1A) by site-directed mutagenesis. Since other parts of the Py enhancer could have responsive elements which may mask the effects of the mutations, we used as a test gene a recombinant construct (pA411) which contains a hybrid promoter (Fig. 1b, and ref. 6) with other essential transcription elements derived from the SV40 early promoter which is relatively insensitive to *ras* activation (C.W., unpublished observations).

The Py enhancer is poorly active in mouse myeloma MPC11-BU4 cells and is strongly activated by *ras* expression and by the tumour promoter TPA³. These myeloma cells were transfected

Received 30 December 1987; accepted 28 January 1988.

- Bishop, J. M. *Cell* **42**, 23–38 (1985).
- Bishop, J. M. & Varmus, H. in *RNA Tumor Viruses, Molecular Biology of Tumor Viruses* Vol. 2 (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 249–356 (Cold Spring Harbor Laboratories, 1985).

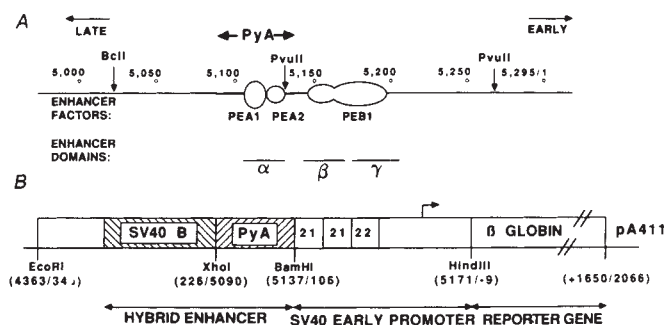


Fig. 1 Principal features of the Py enhancer (A) and structure of test recombinant construct pA411 (B). A, The transcription factors PEA1, PEA2 and PEB1¹³ bind to the Py enhancer, in the functional domains α , β and γ ¹⁴⁻¹⁶. B, pA411 contains Py A enhancer sequences (see part A) in the SV40 early promoter. The hybrid promoter lies upstream from rabbit β globin structural gene sequences⁶.

with pA411, or other mutated derivatives, and after 48 hours the level of transcription was measured by quantitative S1 nuclease analysis of RNA. A co-transfected recombinant construct was always used to control for variations in transfection efficiency. In a large series of experiments, it was reproducibly observed (standard deviation $\sim 10\%$) that point-mutations in sequences from 5097 to 5125 (all mutants besides pA429, Fig. 2) had very little or no effect on transcription from the SV40 early promoter. In contrast pA429 decreased transcription (Fig. 2). Two proteins have been described which bind specifically to the α domain of the Py enhancer, PEA1 and PEA2 (Fig. 1A, ref. 4). The pA429 mutation lies exclusively in the PEA2 binding site, whereas two others (pA418, pA426) alter exclusively the PEA1 binding site, suggesting that, in contrast to PEA2, PEA1 cannot productively interact with the Py α domain in the non-treated myeloma cells. Treatment of transfected cells with TPA activated transcription from the wild-type recombinant (pA411, Fig. 2). Mutations outside the PEA1 binding site had no significant effect on activation (pA421, pA429, Fig. 2). Large mutations in either half of the PEA1 binding site significantly decreased TPA activation (pA418, pA422, Fig. 2). With more

precise mutations it was found that centrally located alterations greatly decreased activation (pA426, pA428, Fig. 2) whereas flanking mutations had a less striking effect (pA425, pA427, Fig. 2). These results show that there is a TPA-responsive element between 5114 and 5121 in the Py enhancer α domain. To test whether the same sequences mediate ras activation, the recombinants were transfected with a ras expression vector or a control expression vector (see Fig. 2 legend and ref. 3). Ras expression stimulated transcription from the wild-type recombinant (pA411, Fig. 2) whereas two precise mutations in the TPA-responsive element greatly diminished ras-induced gene expression (pA426, pA428, Fig. 2). Mutations outside this sequence had no significant effect on activation by ras (pA429, Fig. 2 and other data not shown). These results show that the ras-responsive element (RRE) has the same sequence as the TPA-responsive element and lies in the binding site for PEA1⁴.

The Py enhancer is active in LMTK⁻ fibroblasts grown in high serum (10%) and is not stimulated by ras expression or by TPA³. Transfections of pA411 and its derivatives in these cells showed that mutations in the RRE decreased the activity of the Py α enhancer, and that TPA and ras expression had no effect on transcription from the wild type or RRE mutant recombinants (pA411, pA426, pA428, Fig. 2). These results indicate that RRE, like the Py enhancer, is active in high serum (10%) in these cells and cannot be further stimulated by TPA or ras expression.

Ras protein is involved in signal transduction between growth factors and intracellular targets^{1,2}, suggesting that serum components could act through the RRE sequences. LMTK⁻ cells were transfected in high serum (10%) and then incubated for 24 hours in low serum (0.05%). Transcription from the wild type recombinant decreased dramatically (pA411, Fig. 2), whereas transcription from the RRE mutants was decreased to a smaller extent (pA426, pA428, Fig. 2). In low serum conditions, ras expression and TPA treatment efficiently stimulated transcription from the wild-type recombinant, whereas the RRE mutants were much less sensitive to stimulation (pA411, pA426, pA428). Adding back serum (10%) also stimulated transcription (data not shown). These results demonstrate that in LMTK⁻ cells the RRE defined in myeloma cells mediates transcriptional activation by serum, TPA and ras. They also show that ras protein or TPA can substitute for serum in the activation of the

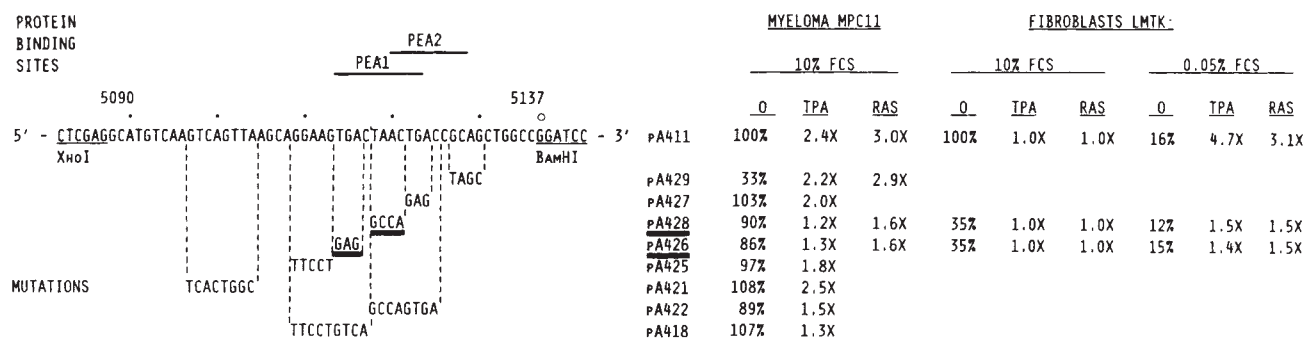


Fig. 2 The effects of Py enhancer mutations on constitutive and ras-, TPA- or serum-induced transcription in mouse myeloma (MPC11-BU4) and fibroblast (LMTK⁻) cells. The binding sites for PEA1 and PEA2 are shown above the sequence of the PyA enhancer (see Fig. 1) and the sequence alterations (recombinants pA4...) are shown below. The results are derived from quantitative analysis by S₁ nuclease mapping of total specific RNA initiated from the SV40 early-early startsites (arrow in Fig. 1B). They summarize a large number of experiments with a minimum of 3 transfections with each of two different DNA preparations. The standard deviations were $\sim 10\%$. The results in the absence of inducer (0) are given as a percentage of wild type (pA411) in high serum and otherwise as the increase on induction for each recombinant. **Methods.** Transfection and quantitative S₁ nuclease mapping were as described previously³. Transfections contained the test recombinant (pA4..., 2 μ g in myeloma cells, 1 μ g in fibroblasts) an equal molar amount of an internal control (pA411) and where appropriate either the ras (pRCM in myeloma cells and pRCBX2 in fibroblasts, ref. 3) or control (pΔRCM and pΔRCBX2 in myeloma cells and fibroblasts, respectively) expression vectors. pA411 contains a deletion between the Klenow repaired HindIII (5171) and AvrII (5187) restriction sites of pA411. pΔRCM and pΔRCBX2 have a deletion of the NarI (1693-1785) fragment, which prevents expression of ras. TPA treatment was for 24 hours, at 100 μ g ml⁻¹. For low serum experiments, LMTK⁻ fibroblasts were transfected in high serum (10%), and then incubated in 0.05% FCS for 24 hours.

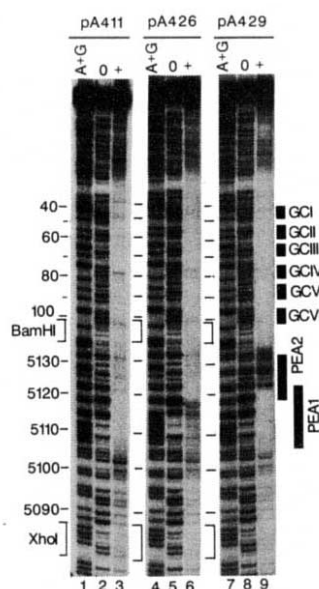


Fig. 3 Footprints of transcription factors in *in vitro* extracts of mouse fibroblast cells. Nuclear extracts from fibroblasts (LMTK⁻) grown in Dulbecco medium with 10% fetal calf-serum were analysed by DNAase I footprinting with 5'-[³²P]-EcoRI (346)-HindIII (5171) probes from pA411 (lanes 2, 3), pA426 (lanes 5, 6) and pA429 (lanes 8, 9) (see Figs 1 and 2) as described in ref. 14. Lanes 1, 4, 7, A+G specific chemical degradation products; lanes 2, 5, 8, without extract; lanes 3, 6, 9, with 20 µg protein. The coordinates and restriction sites at the extremities of the Py enhancer fragment, as well as the binding sites for PEA1, PEA2 and Sp1 (GCI-VI) are indicated.

transcription element. In myeloma cells there may be a block in the signal transduction process from serum to RRE.

To test whether mutations in the RRE prevent PEA1 binding, nuclear extracts were prepared from LMTK⁻ cells grown in high serum (10%) and analysed for the presence of proteins which can specifically protect the promoter region of pA411 from DNase I digestion. The Py enhancer sequences previously reported to bind PEA1 were protected from DNase I digestion compared to control digestions without extract (see nucleotides 5105 to 5153, lanes 2, 3, Fig. 3). A mutation in RRE (pA426, nucleotides 5114 to 5116) prevented PEA1 binding (Fig. 3, lanes 3, 6). In contrast, a mutation in the PEA2 binding site prevented binding of PEA2 but not of PEA1 (pA419, Fig. 3, lanes 8, 9). These results show that mutation of RRE prevents binding of PEA1, suggesting that PEA1 is a transcription factor whose activity is regulated by ras, TPA and serum components.

The complete Py enhancer is about ten times more sensitive to ras protein and TPA activity than the PyA region (see ref. 3 and above). To exclude that this is due to a low sensitivity of RRE to stimulation compared to other potential responsive elements in the Py enhancer, we constructed a recombinant containing multiple PyA enhancer sequences without the other SV40 promoter elements which may have masked stimulation. pG1P4 contains four head to tail copies of an oligonucleotide containing the RRE inducible element located directly upstream from the control β-globin promoter (Fig. 4). Transfections in myeloma MPC11-BU4 cells showed that the multimer alone had little or no effect on globin transcription whereas ras protein expression and TPA treatment stimulated the activity of the multimer to an even greater extent than the activity of the Py enhancer (Fig. 4, lanes 1-8 and ref. 3). Using RRE mutated multimers we have established that RRE mediates this activation (data not shown). These results demonstrate that RRE activity is extremely sensitive to stimulation by ras expression and by TPA in myeloma cells.

This paper describes the identification of a transcription ele-

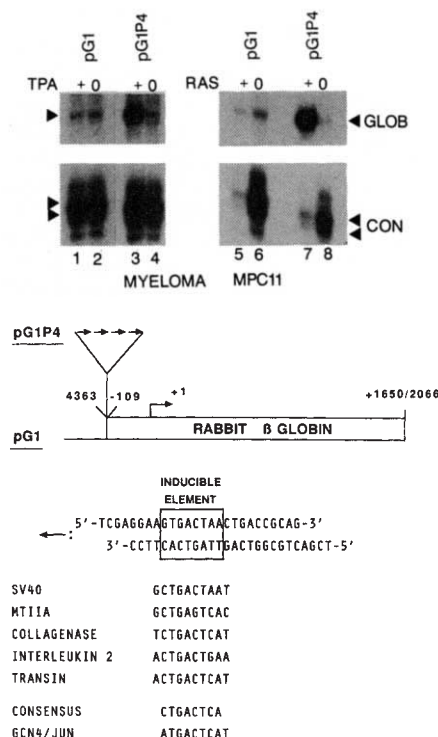


Fig. 4 Ras and TPA activation of a multimer containing the ras-inducible sequence element in mouse myeloma MPC11-BU4 cells. GLOB and CON indicate the specific products obtained by quantitative S1 nuclease mapping for RNA initiated from the globin (pG1P4, pG1) and conalbumin (pRCM, pβCM, ref. 3) promoters, respectively. The sequence of the Py-inducible element is compared with the sequences of the TPA-inducible elements of the SV40 enhancer^{6,9}, the metallothionein MTIIA gene⁹, the collagenase gene¹⁸, the interleukin-2 gene¹⁹, and with similar sequences in the TPA- and ras-inducible promoter from the transin gene¹⁰. GCN4 and the oncogene *jun*^{5,12} recognize a sequence similar to the consensus.

Methods. pG1P4 contains four head to tail copies of the oligonucleotides shown in the Figure (→). They were inserted in the *Xho*I site at -109, directly upstream from the globin promoter in pG1³. The differences in the levels of CON transcription probably result from differences in the stability of RNA transcribed from ras and control expression vectors, and were reproducibly observed (see ref. 3 for evidence).

ment whose activity is regulated by ras, as well as serum and a tumour promoter. Such regulation of a common element supports the hypothesis that ras protein is involved in a signal transduction pathway which involves growth factors and protein kinase C. The recent results of Lacal *et al.*^{7,8}, using DNA replication as an assay, also support this hypothesis. The Py inducible element is remarkably similar to the TPA-inducible elements found in the SV40 enhancer, and the metallothionein MTIIA, collagenase, interleukin-2 and transin genes (ref. 9 and Fig. 4), suggesting that these promoter elements may also be regulated by ras. Indeed, the transformation-induced gene, transin, is activated by ras and contains a RRE (ref. 10, Fig. 4). These results suggest that ras regulates the activity of a number of transformation-related genes through its effect on responsive elements in their promoters, and that an important event in transformation by ras is an uncontrolled constitutive activation of RREs in transformation-related genes.

Mutation of RRE prevents binding of the transcription factor PEA1 (ref. 4), suggesting that the activity of this factor is regulated by ras, TPA and serum. However, we cannot exclude that there are different transcription factors with the same recognition sequence which are differently regulated. RRE is remarkably similar to the sequences recognised by the known enhancer factor AP1 (refs 9, 11, Fig. 4), the oncogene *jun* (refs

5, 12 and M. Karin, personal communication) and a yeast transcription factor GCN4 (ref. 12). These results suggest that PEA1, AP1 and jun may be the same, or closely related proteins, which have been highly conserved during evolution. We have recently demonstrated that *v-jun* activates transcription through RRE (manuscript submitted), suggesting that jun and ras are linked in a pathway which is important in cell transformation.

We thank P. Chambon for useful discussions, M. Zenke for gift of recombinants, P. Chambon, G. Richards, H. Gronemeyer and C. Hart for critical reading of the manuscript, Adrien Staub for the synthesis of oligonucleotides, the staff of the cell culture service for cells and media, C. Werlé and B. Boulay for illustrations, H. Martina for typing the manuscript. This work was supported by grants from the CNRS, the INSERM, the FRM, the MRT and the ARC.

Received 16 October 1987; accepted 5 January 1988.

- Barbacid, M. A. *Rev. Biochem.* **56**, 779-827 (1987).
- Micheal, R. & Jackson, T. *Nature* **328**, 668-669 (1987).
- Wasyluk, C., Immler, J. L., Perez-Mutul, J. & Wasyluk, B. *Cell* **48**, 525-534 (1987).
- Piette, J. & Yaniv, M. *EMBO J.* **6**, 1331-1337 (1987).
- Bohmann, D. *et al. Science* **238**, 1386-1392 (1987).
- Zenke, M. *et al. EMBO J.* **5**, 387-397 (1986).
- Lacal, J. C., Fleming, T. P., Warran, B. S., Blumberg, P. M. & Aaronson, S. A. *Molec. cell. Biol.* **7**, 4146-4149 (1987).
- Lacal, J. C., Moscat, J. & Aaronson, S. A. *Nature* **330**, 269-272 (1987).
- Angel, P. *et al. Cell* **49**, 729-739 (1987).
- Matrisian, L. M., Glaichenhaus, N., Gesnel, M. C. & Breathnach, R. *EMBO J.* **4**, 1435-1440 (1985).
- Lee, W., Mitchell, P. & Tjian, R. *Cell* **49**, 741-752 (1987).
- Struhl, K. *Cell* **50**, 841-846 (1987).
- Kryszke, M. H., Piette, J. & Yaniv, M. *Nature* **328**, 254-256 (1987).
- Veldman, G. M., Lupton, J. & Kamen, R. *Molec. cell. Biol.* **5**, 649-658 (1985).
- Hassel, J. A., Muller, W. J. & Mueller, C. R. in *Cancer Cells*, vol. 4 (eds Botchan, M., Grodzicker, T. & Sharp, P.) (Cold Spring Harbor Laboratory, New York, 1986).
- Herbomel, P., Bourachot, B. & Yaniv, M. *Cell* **39**, 653-662 (1984).
- Wildeman, A. G. *et al. Molec. cell. Biol.* **6**, 2098-2105 (1986).
- Angel, P., Baumann, I., Stein, B., Delius, H., Rahmsdorf, H. J. & Herrlich, P. *Molec. cell. Biol.* **7**, 2256-2266 (1987).
- Fujita, T., Takaoka, C., Matsui, H. & Taniguchi, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7437-7441 (1983).

Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA

Alec J. Jeffreys, Nicola J. Royle, Victoria Wilson & Zilla Wong

Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

Tandem-repetitive minisatellite regions in vertebrate DNA frequently show substantial allelic variation in the number of repeat units. This variation is thought to arise through processes such as unequal crossover or replication slippage¹⁻¹². We show here that the spontaneous mutation rate to new length alleles at extremely variable human minisatellites is sufficiently high to be directly measurable in human pedigrees. The mutation rate at different loci increases with variability in accord with the neutral mutation/random drift hypothesis, and rises to 5% per gamete for the most unstable human minisatellite isolated. Mutations are sporadic, occur with similar frequencies in sperm and oocytes, and can involve the gain or loss of substantial numbers of repeat units, consistent with length changes arising primarily by unequal exchange at meiosis. Germline instability must therefore be taken into account when using hypervariable loci as genetic markers, particularly in pedigree analysis and parenthood testing.

Five different human minisatellites (pAg3, λ MS1, λ MS8, λ MS31 and λ MS43) which have been cloned from DNA fingerprints^{8,12} were chosen for study. Each detects a single hypervariable locus with multiple length alleles and heterozygosities ranging from 85% for λ MS8 to ~99% for λ MS1. To determine

mutation rates to new length alleles, each minisatellite was hybridized to *HinfI* digests of DNA from the Centre d'Etude du Polymorphisme Humain (CEPH) panel of 40 large families, giving a total of 344 offspring (688 gametes) who could be scored for new mutation. The correct parentage of all individuals analysed has been established by extensive typing with conventional markers and restriction fragment length polymorphisms (RFLPs) (J. Dausset and H. Cann, personal communication) and confirmed with minisatellite probes (data not shown).

Examples of new mutations detected are shown in Fig. 1. These mutations could be detected using any restriction endonuclease which does not cleave within the minisatellite (data not shown), and therefore result from a change in the length, and presumably number of repeat units, of a minisatellite. All mutant offspring appeared to be clonal, with no trace of the original parental allele in addition to the new mutant allele. The lack of mosaicism expected for somatic changes in polyclonal blood cells or lymphoblastoid cell lines from which the CEPH DNAs were prepared suggests that mutations arise in the germline or possibly in the early embryo in a single precursor cell of all B-lymphoblasts. Furthermore, minisatellite length changes were heritable in six families where the mutation had arisen on transmission from grandparent to parent; in all cases the new mutant allele was faithfully transmitted from parent to offspring (Fig. 1c). We therefore conclude that most or all mutations observed are germline in origin.

Mutations arise sporadically and showed no significant clustering within families; for example, the 36 mutant alleles detected by λ MS1 were distributed over 22 out of 40 families, compared with 24 out of 40 families expected for a random distribution of mutant offspring. Furthermore, different offspring in a sibship never shared a common mutant allele (Fig. 1), indicating that there is no significant level of germline mosaicism for mutant alleles, in contrast to recent evidence for germline mosaicism of Duchenne muscular dystrophy mutations^{13,14}. This lack of mosaicism is consistent with new mutants being generated preferentially in the latest stages of gametogenesis, although it remains possible that mutations might arise with equal frequency at all stages of development of the germline (analysis not shown).

The mutation rate to new length alleles varies substantially from locus to locus, and rises to 5.2% per gamete for the most unstable locus λ MS1 (Table 1). The mutation rate increases with heterozygosity (Table 1) and the relationship between mutation rate and heterozygosity is in accord with the theoretical predictions of the neutral mutation/random drift hypothesis (Fig. 2). Two models of allelic change (small stepwise mutations¹⁵ and random mutation of an allele to any other allelic state¹⁶) predict effective population sizes N_e of ~50,000 and ~3,000 respectively. Since the process of allelic length change is intermediate between these two models (see below), N_e should lie between these two estimates, in reasonable agreement with an estimate of ~5,000 for most large, national populations derived from demographic analysis¹⁷. Neutral mutation rates therefore appear to be sufficiently high to maintain the level of variability seen at these hypervariable loci.

Length changes at λ MS1 involve the gain or loss of variable numbers (4-200) of nine base pair (bp) repeat units (Fig. 3). It would not be possible to resolve smaller length changes by electrophoresis, and their incidence is unknown. There is no clear relationship between the size of the mutation and the length of the allele (Fig. 3b), but there is (weak) statistical evidence suggesting that mutation rate is proportional to allele length (Fig. 3a and 3b, analysis not shown). Assuming proportionality and taking into account the resolution limits of the gels (Fig. 3b), we estimate that ~50% of mutation events involving λ MS1 allele length changes of at least four repeat units would not have been resolved in this screen. The mutation rate at λ MS1 may therefore be as high as ~10% per gamete.

The incidence of mutations involving gain or loss of repeat

Fig. 1 New length variants of minisatellites detectable in three human pedigrees (*a*, *b* and *c*). Squares, males; circles, females. Alleles within each family are identified by letters and mutant alleles are arrowed. Sizes of alleles in kilobases (kb). Since most mutations result in a relatively small DNA fragment length change, the progenitor of each mutant allele (asterisked) could be tentatively identified as the parental allele most similar in size to the mutant allele. For example, the mutant allele of λ MS31 in the family shown in (*a*) is maternal in origin and is most probably derived by contraction of allele D rather than C. This assignment has been confirmed by retesting this family with *Hae*III, which cleaves several times within λ MS31 to produce a haplotype of linked minisatellite DNA fragments from each allele; as predicted, the mutant allele shares DNA fragments with allele D but not C (data not shown). It is not possible to determine whether allele length change has arisen by unequal exchange between parental alleles at meiosis; if this occurs, then mutant alleles would be hybrids of the two parental alleles. Family (*b*) contains several independent mutations at λ MS1, including one double-mutant child. The correct parentage of this child has been confirmed using probes $p\lambda$ g3, λ MS8, λ MS31 and λ MS43 and by DNA fingerprinting. The mutation, probably in allele A, in the third child is close to the gel resolution limit, with a size decrease of ~ 40 bp ($\sim$ four 9 bp repeat units). In family (*c*), an allele (probably B) in the paternal grandmother has mutated on transmission to the father. The mutant allele B* is heritable in that it is stably transmitted to the father's four children who have not inherited paternal allele A.

Methods. 2 μ g samples of DNA from Epstein Barr virus-transformed lymphoblastoid cell lines from the CEPH panel of families were digested with *Hinf*I and electrophoresed through a 20 cm long 0.8% agarose gel until all DNA fragments < 1 kb had run off the gel. Narrow (1.5 mm) loading slots were used to assist allele mobility comparisons across families. After electrophoresis, DNA fragments were denatured, transferred to Hybond-N (Amersham) and hybridized with 32 P-labelled minisatellite probes λ MS1 or λ MS31 as described^{1,2}.

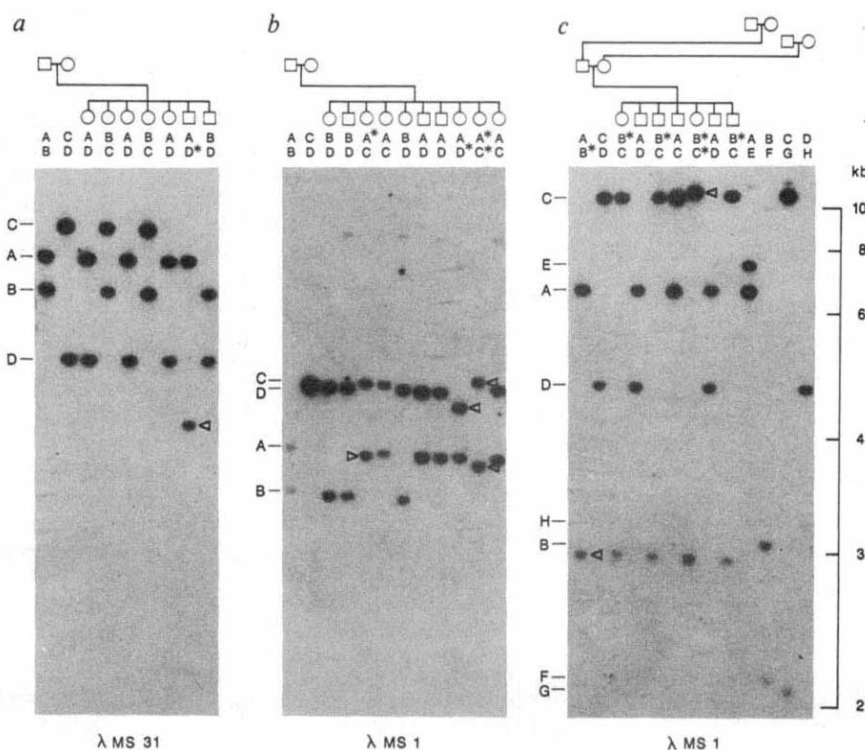


Table 1 Summary of heterozygosities and mutation rates at five human minisatellite loci

Probe	Locus	Repeat length bp	per cent heterozygosity		Mutation rate per gamete	Number of mutant alleles observed					
			a	b		Paternal			Maternal		
						i	d	u	i	d	u
λ MS8	<i>D5S43</i>	29, 30	89	85.1 (82.8–87.2)	0 (0–0.003)	0	0	0	0	0	0
λ MS43	<i>D12S11</i>	45	97	95.9 (94.7–96.9)	0 (0–0.003)	0	0	0	0	0	0
$p\lambda$ g3	<i>D7S22</i>	37	98	97.4 (96.4–98.1)	0.003 (0.0006–0.009)	1	0	0	1	0	0
λ MS31	<i>D7S21</i>	20	98	98.0 (97.1–98.6)	0.007 (0.003–0.015)	3	0	0	0	2	0
λ MS1	<i>D1S7</i>	9	100	99.4 (98.9–99.7)	0.052 (0.038–0.072)	9	9	1	7	7	3
Total						13	9	1	8	9	3
Total increased							23			20	
Total decreased										21	18

Restriction fragment length mutations were characterized in 40 large families from the CEPH panel comprising a total of 99 grandparents, 80 parents and 296 children. DNA samples from all individuals were digested with *Hinf*I and hybridized sequentially with five human minisatellite probes as in Fig. 1. Heterozygosities (*a*) for each locus in these families were determined by counting the number of resolvable heterozygotes amongst 125 unrelated grandparents or, where unavailable, parents. Heterozygosities (*b*) were more accurately estimated from the degree of allele sharing observed between unrelated individuals^{1,2}, based on 350 independent pairwise comparisons of unrelated people from the CEPH panel. Heterozygosities estimated from comparisons between families were not significantly different from estimates restricted to comparisons between grandparents or parents within individual families (data not shown). Mutations were scored in a total of 344 offspring (688 gametes); all mutations resulting in small changes in DNA fragment length were confirmed by retesting the mutant child alongside the parents. 90% confidence limits are given for heterozygosities (*b*) and for mutation rates. Heterozygosities and mutation rates are a mean for the population groups represented in the CEPH panel (27 Utah families, 10 French, 2 Venezuelan and 1 Amish); more detailed analysis showed no significant variation in either mutation rate or heterozygosity between these limited samples of different population groups. The number of mutant paternal and maternal alleles observed which appeared to have increased in size (*i*) or decreased in size (*d*) are given; in some instances where the size of the new mutant allele was intermediate between the sizes of the two parental alleles, it was not possible to estimate the direction of mutation (*u*).

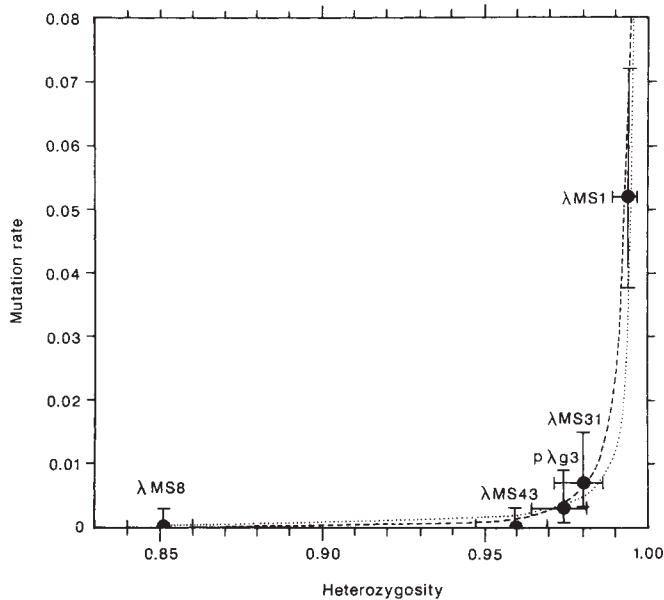


Fig. 2 Relationship between heterozygosity and mutation rate at five hypervariable minisatellite loci. Estimates of mean heterozygosity of populations within the CEPH panel and mutation rates per locus per gamete, together with 90% confidence limits, are taken from Table 1. ----, best-fit theoretical curve predicted by the neutral mutation/random drift hypothesis under the stepwise mutation model, in which each mutation increases or decreases the length of an allele by the minimum amount resolvable by gel electrophoresis. In this model, homozygosity H_0 and mutation rate u are related by $H_0 = 1/(1 + 8N_e u)^{1/2}$ where N_e is the effective population size¹⁵; the best-fit curve is given with $N_e = 50,000$, best-fit curve predicted under the symmetric K allele model in which mutations occur in equal frequencies in all directions among K alleles. Assuming selective neutrality, $H_0 = [1 + 4N_e u/(K - 1)]/[1 + 4N_e uK/(K - 1)]$ (ref. 16). K can be estimated from the resolution limit of the gels; DNA fragments > 1 kb were resolved over a distance of 150 mm, and the minimal resolvable change in electrophoretic migration is ~ 0.5 mm. There are therefore $150/0.5 = 300$ resolvable allelic states, giving $K = 300$. The best-fit curve under the K allele model is obtained with $N_e = 3000$. The difference in estimates of N_e from the stepwise mutation and K allele models reflects the tendency of step-wise mutations to produce, by back mutation, pre-existing alleles; the number of different alleles maintained in a population by a given mutation rate is therefore lower with step-wise mutation than with random mutation.

units at λ MS1 are similar (Table 1). On average, 38 ± 10 (s.e.m.) repeat units appear to be gained per mutation event, compared with 30 ± 13 units lost per event, and corresponding to a mean change in allele repeat copy number of 5% (Fig. 3). There is therefore no substantial bias towards expansion of λ MS1 alleles, suggesting that the largest alleles accumulate fortuitously through genetic drift.

The mutation rate at λ MS1 is indistinguishable in sperm and oocytes (Table 1). This is surprising since a female's complement of oocytes arise through ~ 24 postzygotic cell divisions whereas ~ 400 divisions separate the male zygote from mature sperm¹⁸. It therefore appears that most of these length changes are not replication- or cell division-dependent, consistent with the absence of germline mosaicism and arguing against mitotic recombination or replication slippage being predominant mutation processes. However, the size distribution of mutations differs significantly between sperm and oocytes, with sperm showing an excess of small (4–10 repeat) mutation events (Fig. 3c). The small size of these events and the relatively high rate in sperm is consistent with their generation through replication slippage events¹⁹. The remaining large mutation events, involving the gain or loss of up to 200 repeat units, appear to be

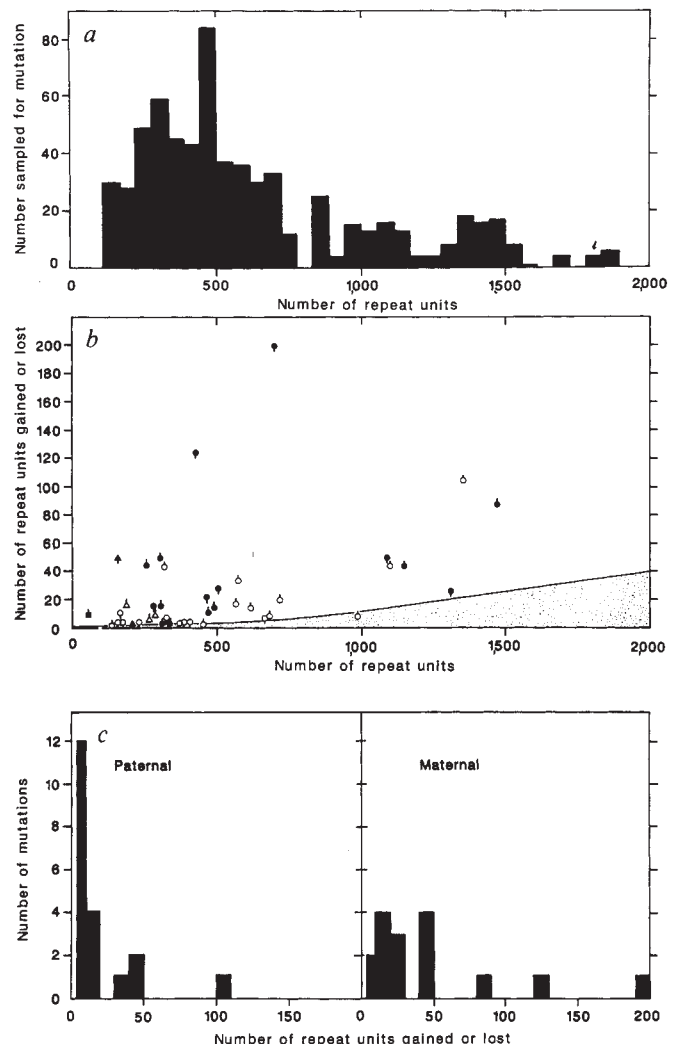


Fig. 3 Characteristics of mutations detected by minisatellite probe λ MS1. *a*, Size distribution of parental λ MS1 alleles sampled for mutation in the CEPH families; since λ MS1 *Hinf*I DNA fragments contain 490 bp flanking DNA in addition to the minisatellite, the number of nine bp repeat units can be determined from *Hinf*I fragment length. The allele size distribution is given in 500 bp (56 repeat) size intervals, though each size interval contains multiple resolvable alleles. *b*, Relationship between the estimated number of repeat units gained or lost in mutant λ MS1 alleles and the number of repeat units in the (provisionally identified) parental allele. $\circ$, mutant paternal allele with gain of repeats; $\square$, mutant paternal allele with loss of repeats; $\bullet$, mutant maternal allele with gain of repeats; $\blacksquare$, mutant maternal allele with loss of repeats respectively. The shaded area represents the approximate gel resolution limit; any λ MS1 mutations within this area will result in < 0.5 mm change in electrophoretic mobility, and will not be resolvable from the parental allele. For comparison, mutations detected by λ MS31 ($\triangle$ = paternal, $\blacktriangle$ = maternal) and $p\lambda g3$ ($\square$ = paternal, $\blacksquare$ = maternal) are also shown. *c*, Frequency distributions, in 10 repeat unit intervals, of size changes in mutant paternal and maternal alleles detected by λ MS1. These frequency distributions differ significantly ($P < 0.05$, two-sample two-tailed Kolmogorov-Smirnov test²²).

replication-independent and might arise by recombinational processes such as unequal exchange or gene conversion at meiosis. This is consistent with earlier speculations that the 'core' sequence associated with many minisatellites⁴, including λ MS1 and $p\lambda g3$ ^{8,12}, might serve as a recombination signal in human and animal DNA^{4,9}. A more detailed analysis of mutant minisatellite structure and of flanking markers should clarify the possible role of meiotic recombination in the generation of new length alleles of minisatellites.

The mean mutation rate to new length alleles in DNA fingerprints detected by multi-locus polycore probes has been estimated at ~ 0.004 per DNA fragment per gamete (ref. 4, I. Patel and A.J.J., unpublished data). λ MS1 must therefore be one of the most unstable loci detected in DNA fingerprints. The use of DNA fingerprints to establish family relationships^{5,9,10,20} is not significantly impeded by the occasional new mutation, in view of the large number of additional informative markers which will confirm or deny the claimed relationship. However, parentage testing with a limited number of hypervariable locus-specific probes^{12,21} could lead to the false exclusion of genuine parents if mutation rates are significant. Direct measurement of spontaneous mutation rates, or indirect estimation from heterozygosity via the theoretical curves shown in Fig. 2, is a necessary prerequisite for reliable determination of parentage.

We thank Professor J. Dausset and H. Cann for providing the panel of CEPH DNAs and Dr R. Semeonoff for helpful discussions. This work was supported by a grant from the MRC. A.J.J. is a Lister Institute Research Fellow. The minisatellite probes are the subject of patent applications, and commercial enquiries should be addressed to ICI Diagnostics, Gadbrook Park, Rudheath, Northwich, Cheshire, CW9 7RA, UK.

Received 19 November 1987; accepted 18 February 1988

1. Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31–35 (1982).
2. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33–37 (1983).
3. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5022–5026 (1983).
4. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73 (1985).
5. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76–79 (1985).
6. Jarman, A., Nicholls, R. D., Weatherall, D. J., Clegg, J. B. & Higgs, D. R. *EMBO J.* **5**, 1857–1863 (1986).
7. Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. & Ponder, B. A. J. *Am. J. hum. Genet.* **39**, 11–24 (1986).
8. Wong, Z., Wilson, V., Jeffreys, A. J. & Thein, S. L. *Nucleic Acids Res.* **14**, 4605–4616 (1986).
9. Jeffreys, A. J. *Biochem. Soc. Trans.* **15**, 309–317 (1987).
10. Jeffreys, A. J. & Morton, D. B. *Anim. Genet.* **18**, 1–15 (1987).
11. Nakamura, Y. *et al Science* **235**, 1616–1622 (1987).
12. Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. J. *Ann. hum. Genet.* **51**, 269–288 (1987).
13. Bakker, E. *et al Nature* **329**, 554–556 (1987).
14. Darras, B. T. & Francke, U. *Nature* **329**, 556–558 (1987).
15. Ohta, T. & Kimura, M. *Genet. Res.* **22**, 201–204 (1973).
16. Kimura, M. *Genet. Res.* **11**, 247–269 (1968).
17. Morton, N. E. *Outline of Genetics Epidemiology* (Karger, Basel, 1982).
18. Vogel, F. & Rathenberg, R. *Adv. hum. Genet.* **5**, 223–318 (1975).
19. Levinson, G. & Gutman, G. A. *Molec. Biol. Evol.* **4**, 203–221 (1987).
20. Jeffreys, A. J., Brookfield, J. F. Y. & Semeonoff, R. *Nature* **317**, 818–819 (1985).
21. Baird, M. *et al Am. J. hum. Genet.* **39**, 489–501 (1986).
22. Siegel, S. *Nonparametric Statistics for the Behavioural sciences* (McGraw-Hill, Tokyo, 1956).

Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres

Diethard Tautz

Max-Planck Institut für Entwicklungsbiologie, Arbeitsgruppe Jäckle, Spemannstrasse 35/II, 7400 Tübingen, FRG

Segmentation in the insect embryo is initiated by maternally provided information, which is stored in the developing oocyte^{1–6}. In *Drosophila*, the genes necessary for this process have been genetically characterized^{2–6}. The anterior segmented region is organized by the *bicoid* (*bcd*) gene product². The posterior segmented region is organized by several interacting gene products, among them the *oskar* (*osk*)^{3–5} gene product. The first zygotic group of genes, which are thought to respond to the spatial cues provided by the maternal genes, are the gap genes^{7–9}, whose members include *hunchback* (*hb*), *Krüppel* (*Kr*) and *knirps* (*kni*)⁷. To elucidate the role played by the maternal genes in expression of the gap gene *hb*, antibodies were raised against a fusion protein and were used for the cytological localization of the *hb* gene product in wild-type and mutant embryos. The *hb* protein is pre-

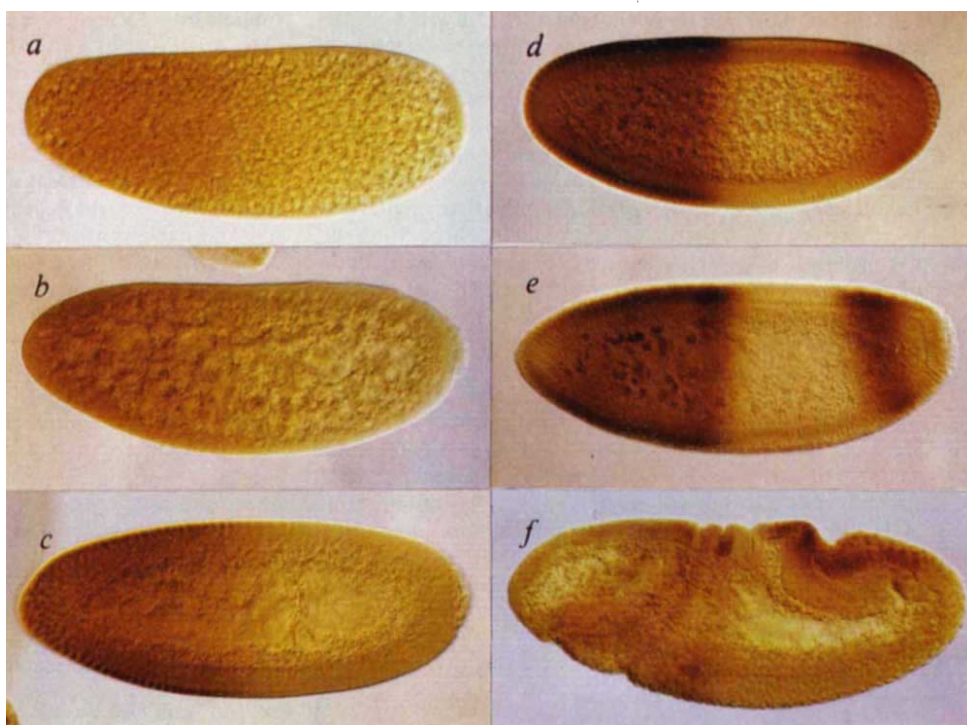
dominantly located in the nucleus. Its spatial expression includes the formation of an anterior-posterior gradient during the early cleavage stages and a strong zygotic expression in the anterior half of the embryo. Analysis of embryos mutant for the maternal genes affecting the anterior-posterior segmentation pattern shows that the formation of the early gradient is controlled by the *osk* group of genes, whereas efficient activation of the zygotic anterior expression domain is dependent on *bcd* activity.

The gap segmentation genes are thought to provide a first subdivision of the embryo, which is further interpreted by the pair rule and the segment polarity genes to form eventually the segmented pattern of the larvae^{7–9}. The *hb* gene takes a special place among the gap genes, as its product is provided not only zygotically, but also maternally. Embryos mutant for *hb* show a stronger phenotypic effect, if the maternal gene product is missing. On the other hand, embryos which lack the maternal gene product can still be rescued, if a wild-type copy of the gene is provided paternally¹⁰. *In situ* hybridization to tissue sections of early embryos has shown that the maternally provided *hb* RNA forms an anterior-posterior gradient during the early cleavage stages¹¹. At blastoderm stage, *hb* is expressed in an anterior and a posterior domain¹¹ which both reflect the phenotypic effects of *hb*^{10,12}, although the posterior domain appears to be under independent control¹¹.

The complete open reading frame of the *hb* coding region was used to produce a fusion protein in *Escherichia coli*, which was used to immunize a rabbit (see Fig. 1 for details). The purified serum was used to stain the *hb* protein in early wild-type embryos up until gastrulation (Fig. 1). Freshly-laid eggs show no staining, implying that no *hb* protein is delivered by the mother, though *in situ* hybridization has shown that the *hb* RNA is present at this stage¹¹. The first translation of the maternal *hb* messenger RNA is observed before pole-cell formation, before stage 8 (staging refers to nuclear division cycles¹³), when the protein distribution forms a concentration gradient from anterior to posterior (Fig. 1a). Direct densitometric scanning reveals that this gradient is approximately linear over the posterior two thirds of the embryo, although it reaches a plateau in the anterior region (Fig. 2). On progression to stage 10, the protein becomes more and more confined to the anterior half of the embryo (Fig. 1b). Full establishment of the anterior expression domain begins during stages 11–12 (Fig. 1c), coinciding with the zygotic expression of *hb* RNA¹¹. Staining in this domain becomes much stronger during stages 13 and 14 and eventually a fairly clear border at about 51% egg length (EL, posterior is 0%) is established (Fig. 1d, e). At late-stage 14 the staining recedes slightly from the anterior tip of the embryo (Fig. 1e). During gastrulation, *hb* protein expression becomes very weak in the anterior domain and eventually ceases (Fig. 1f). Protein expression in the posterior *hb* domain appears first at mid-stage 14, where it forms a cap around the posterior pole, excluding the pole cells (Fig. 1d). It then recedes from the pole and forms a ring between about 10–20% EL (Fig. 1e). Expression in this domain remains until the end of germband extension. Both the anterior and the posterior domains show an asymmetry in the staining intensity of dorsal and ventral regions, with the dorsal staining being stronger. A qualitative difference between the anterior and the posterior domains is the staining of the yolk nuclei, which are only visible in the anterior domain (Fig. 1e). The *hb* protein staining is predominantly nuclear, which is in line with the suggestion that it might act as a transcription factor¹¹.

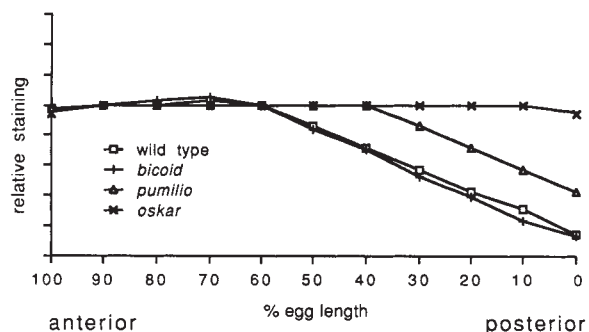
Three groups of maternal genes are involved in the establishment of the anterior-posterior segmented pattern of the embryo and the establishment of the unsegmented terminal regions (reviewed in ref. 6). These are *bcd*, which is the effector gene for the establishment for the anterior segmented pattern², the *osk* group of genes, which establish the posterior segmented pattern^{3–5} and the *torso* (*tor*) class of genes, which organize the terminal regions^{3,6}.

Fig. 1 Wild-type staining pattern of the *hb* protein in early embryos. Whole embryos were fixed with paraformaldehyde, reacted with the *hb* antiserum and stained with a peroxidase detection procedure as described in ref. 15. Optical sections of whole mount embryos are shown. The orientation is always anterior-left and dorsal-up. Staging refers to nuclear division cycles as described in ref. 13. In the embryo before stage 8(a), the pole cells are not yet formed but the gradient is apparent by stage 10(b), pole cells have formed. In the embryo at stage 12(c), *hb* expression is confined to the anterior half. At mid stage 14(d), the posterior domain begins to form. An embryo at late stage 14(e) shortly before the onset of gastrulation, shows retraction of the posterior domain and a shift of the border of the anterior expression domain from about 55% to 51% EL (also seen in the *in situ* hybridization experiments¹¹). Double staining experiments with the *Kr* antibody (U. Gaul, unpublished data) also detect this shift which is not discussed further here, as it appears to be under the control of later acting genes (unpublished data). The three striped and the two striped modulation of the anterior *hb* expression domain which was detected in the *in situ* hybridization experiments¹¹ at this stage, is not visible here but can also be seen in the antibody staining pattern (depending on the staining conditions). An embryo during germband extension is shown in (f). The *hb* alleles *hb*^{7M}, *hb*^{9Q} and *hb*^{F82}, which were characterized as amorphic by genetic criteria (class-I alleles, ref. 10), did not stain with the purified antibody. Two other class-I alleles, namely *hb*^{6N} and *hb*^{14F} showed reduced staining.



Methods. Construction of the *hb* fusion proteins. A 2.4 Kilobase *Xba*I fragment, which covers the entire open reading frame of the *hb* protein (positions 4,670-7,105 of the sequence given in ref. 11), was cloned in frame into the *Xba*I site of the pPTH-11 vector, which contains the start of the *trpE* operon of *E. coli* as an inducible promoter. Induction and isolation of the protein was done as described in ref. 17. The same fragment was used for cloning into the *Xba*I site of the pUR-288 vector, which contains the *lacZ* gene of *E. coli* as the inducible promoter system¹⁸. This protein was used for affinity purification of the antiserum as described in ref. 17.

Fig. 2 Direct densitometric evaluation of the early *hb* gradient in wild-type and mutant embryos. Embryos between stage 6-8 (before pole cell formation) were stained using the glucose oxidase staining procedure (Vectastain, Vector Laboratories), and scanned with a Zeiss single-beam cytophotometer along the anterior-posterior axis. Scanning was done in 5 μ m steps and the transmission values lay between 10 and 80%, which is within the range of linearity. The curves represent averages from several embryos and are normalized to an arbitrary scale to make them comparable with each other. But this method does not allow a direct quantitation of the amount of protein and it is therefore only the slopes of the curves which can be compared and not the absolute values of the staining intensities.

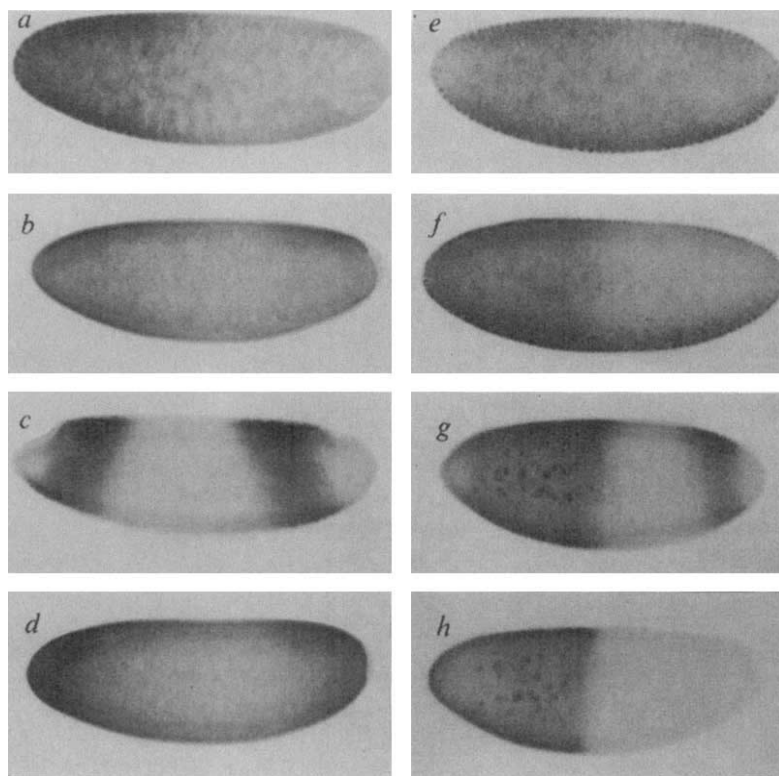


Staining embryos from mothers mutant for strong *bcd* alleles shows that the anterior-posterior *hb* protein gradient forms normally (Fig. 2) and that the protein expression retracts into the anterior domain during stages 11-12 (Fig. 3a). But the strong zygotic activation of *hb* appears to be absent in *bcd* mutant embryos, as there is only weak staining of *hb* protein during stage 13 and early stage 14 (Fig. 3b). This residual staining could be due to the remaining maternally provided *hb* product, or it could be a weak zygotic activation which might be driven by an as yet unknown gene or by autoregulation of the *hb* product alone. These experiments do not distinguish between these possibilities. At mid stage 14, a posterior *hb* domain forms, which is enlarged towards the anterior region (Fig. 3b, c). This is in line with the shifting of the fate map in embryos mutant for *bcd*¹⁴. Also at mid-stage 14, a new domain appears in the anterior region, which is apparently a mirror image duplication of the posterior domain (Fig. 3c). The formation of this new *hb* domain can be correlated with the duplication of posterior

structures in strong *bcd* alleles, resulting in a second 'posterior' midgut invagination². Weak *bcd* alleles which lack the duplication of posterior structures, show the formation of the anterior *hb* domain, though it is reduced in its size, depending on the strength of the *bcd* allele. Interestingly, the weak and *bcd* independent staining of the anterior half of the embryo can still be seen (Fig. 3d). The *bcd* control on *hb* is at the transcriptional level, as *in situ* hybridization to *hb* RNA in strong *bcd*⁻ embryos showed the same result as the protein staining (data not shown).

Staining embryos from mothers mutant for *osk* shows that the normal anterior-posterior gradient of the maternal *hb* product is not formed (Fig. 2). At stage 11-12, all nuclei are stained in *osk*⁻ embryos (Fig. 3e). The degree of staining resembles the weak anterior staining in *bcd*⁻ embryos (see above), though it is now extended to all nuclei. From stage 12-13 on, the anterior *hb* domain begins to form normally and staining of nuclei in the posterior half of the embryo is reduced (Fig. 3f). At late-stage 14, a relatively normal *hb* expression pattern emerges. But the

Fig. 3 *hb* protein staining of embryos mutant for maternal gene products. Embryos were collected from mothers homozygous mutant or transheterozygous mutant for the appropriate genes. Fixation and staining was as for wild-type embryos. *a, b, c*, embryos mutant for *bcd* allele *bcd*^{E1} over deficiency *Df(3R)LIN* (ref. 2). At stage 11–12 (*a*) *bcd*-independent staining is seen in the anterior region. At early stage 14 (*b*), weak anterior staining is still apparent and the posterior domain begins to form. An embryo at late stage 14 (*c*), shows duplicated expression of the posterior *hb* domain. Staining of the yolk nuclei is not apparent in the anterior region, which further supports the notion of a duplicated posterior domain. The same staining pattern was observed for the following *bcd* alleles: *bcd*⁰⁸⁵ and *bcd*^{33–5} over deficiency *Df(3R)LIN*, *bcd*^{E1}, *bcd*^{E85} and *bcd*^{33–5}, homozygous. *d*, embryo mutant for *bcd*^{E4} over deficiency *Df(3R)LIN*, at stage 13–14. Weak staining is seen in the anterior half and strong, *bcd*-dependent staining at the anterior tip. *bcd* alleles which similarly show no duplication of the posterior *hb* stripe, but a reduction in the size of the anterior *hb* domain are, in the order of the severity of the effect: *bcd*^{E4} over deficiency *Df(3R)LIN* (as shown in *d*), *bcd*^{E4} and *bcd*^{E5} homozygous (all alleles are described in ref. 2). *e, f, g*, embryos homozygous mutant for the *osk* allele *osk*¹⁶⁶ (ref. 4). An embryo at stage 12 (*e*), shows staining of all nuclei (note that there are no pole cells formed in *osk*[−] embryos⁴). At early stage 14 (*f*) a normal anterior domain is seen, whereas the nuclei in the posterior region still appear weakly stained. At late stage 14 (*g*) the *hb* expression pattern is relatively normal though the border of the anterior domain is shifted slightly towards the posterior and lies at about 49% EL. A similar shift is observed in embryos mutant for *vasa*^{PD23} and *pumilio*⁶⁸⁰. The shift towards the posterior is more pronounced in *nanos*^{L7}, and extends to about 40% EL. In *staufer*^{D3}, the border of the anterior domain is shifted towards the anterior to about 56% EL. The posterior domain is slightly enlarged towards the anterior in *pumilio* and *nanos* and slightly reduced and shifted towards the posterior in *staufer*. All these effects are somewhat variable, as is to be expected from the partially variable phenotypes of these mutants. *h*, embryo mutant for *torso* at late stage 14.



border of the anterior domain is slightly shifted towards the posterior in *osk*[−] embryos (Fig. 3g). Mutations in three other genes from the *osk* group, namely *vasa*, *staufer*^{3,6} and *nanos* (R. Lehmann, unpublished data) also prevent the formation of the early *hb* gradient, though they show slightly different effects on the size of the anterior *hb* domain and the size and location of the posterior domain (see legend to Fig. 3). Mutations of *pumilio*, another gene of the *osk* group⁵, showed a weaker effect on the formation of the early *hb* gradient. In these embryos, the early *hb* expression covers the anterior two thirds of the embryo and falls off in the posterior third in a gradient-like way (Fig. 2). It is known that mutations in the *pumilio* gene product do not affect the activity, but only the spatial localization of the posterior morphogenetic factors⁵. The *pumilio* effects on *hb* expression are therefore in line with the proposed function of the *pumilio* gene product, namely an involvement in the transport of the active posterior factors towards more anterior positions⁵.

The mechanism of action of the posterior morphogenetic factors on *hb* expression could be a destabilization of the *hb* gene products, but is more likely to be negative translational control of the maternally provided *hb* RNA. This is indicated by the fact inferred from comparison of the protein-staining and *in situ* hybridization¹¹ data that this RNA is homogeneously distributed in freshly laid eggs¹¹ and that its distribution becomes graded only after the protein gradient is apparent. The RNA gradient could be explained by preferential degradation of non-translated RNA, a mechanism which has also been proposed for the formation of the gradient of the *caudal* gene products^{15,16}. The overall phenomenology of the formation of the *hb* RNA

and protein gradients are similar in time, but reciprocal in space to those of the *caudal* gene products. Whether this points to a functional similarity is as yet unclear.

In embryos coming from mothers homozygous mutant for the *tor* gene product, neither the early *hb* gradient, nor the initial zygotic expression in the anterior half of the embryo are affected. But at late stage 14, the posterior *hb* domain is not formed and the retraction of the anterior *hb* domain from the tip of the embryo is not seen (Fig. 3h). These are effects on late *hb* expression and it appears therefore possible that they are related to fate-map changes in the late stage 14 embryo, though a direct regulatory influence can not be ruled out.

My results show that the early expression of *hb* is antagonistically regulated by both morphogenetic centres, though the regulation can be separated into the effects on the maternally provided *hb* RNA and the efficiency of zygotic activation. While the zygotic *hb* activation is obviously necessary for the formation of anterior segments, it is not clear which phenotypic effects may be derived from the maternal expression of *hb*. It has been shown that this expression is not necessary for the survival of the embryos¹⁰, but more subtle effects are not excluded. Our recent experiments do indeed indicate that there may be an *hb*-dependent maternal influence on the correct positioning of the expression domain of *Kr*, the neighbouring gap gene (M. Hülkamp, unpublished data).

I thank H. Jäckle for constant support, D. Sperlich (University of Tübingen) for providing the cytophotometer, U. Gaul for help with the immunization, A. Kienlin for technical assistance and C. Nüsslein-Volhard, R. Lehmann, H.-G. Frohnhofer, M. Hülkamp and C. Schröder for mutant stocks, stained embryos,

discussion and communication of results before publication. This work was supported by the Deutsche Forschungsgemeinschaft (Leibniz Programm to H. Jäckle).

Received 17 August 1987; accepted 15 January 1988.

1. Sander, K. *Adv. insect Physiol.* **12**, 125-238 (1976).
2. Frohnhofer, H. G. & Nüsslein-Volhard, C. *Nature* **324**, 120-125 (1986).
3. Schüpbach, T. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **195**, 302-317 (1986).
4. Lehmann, R. & Nüsslein-Volhard, C. *Cell* **47**, 141-152 (1986).
5. Lehmann, R. & Nüsslein-Volhard, C. *Nature* **329**, 167-170 (1987).
6. Nüsslein-Volhard, C., Frohnhofer, H.-G. & Lehmann, R. *Science* **238**, 1675-1681 (1987).
7. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
8. Meinhardt, H. J. *Cell. Sci. suppl.* **4**, 357-381 (1986).
9. Jäckle, H., Tautz, D., Schuh, R., Seifert, E. & Lehmann, R. *Nature* **324**, 668-670 (1986).
10. Lehmann, R. & Nüsslein-Volhard, C. *Dev. Biol.* **119**, 402-417 (1987).
11. Tautz, D. *et al Nature* **327**, 383-389 (1987).
12. Bender, M., Turner, F. R. & Kaufmann, T. *Dev. Biol.* **119**, 418-432 (1987).
13. Foe, V. E. & Alberts, B. M. *J. Cell Sci.* **61**, 31-70 (1983).
14. Frohnhofer, H. G. & Nüsslein-Volhard, C. *Genes & Dev.* **1**, 880-890 (1987).
15. Macdonald, P. & Struhl, G. *Nature* **324**, 537-545 (1986).
16. Mlodzik, M. & Gehring, W. J. *Cell* **48**, 465-478 (1987).
17. Gaul, U., Seifert, E., Schuh, R. & Jäckle, H. *Cell* **50**, 639-647 (1987).
18. Rüther, U. & Müller-Hill, B. *EMBO J.* **2**, 1791-1794 (1983).

Zinc-finger motifs expressed in *E. coli* and folded *in vitro* direct specific binding to DNA

Kiyoshi Nagai, Yukinobu Nakaseko, Kim Nasmyth & Daniela Rhodes

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The short sequence motif named 'zinc finger', first recognized repeated in tandem in the *Xenopus* transcription factor IIIA (TFIIIA)^{1,2}, is also found in the yeast transcriptional activator SWI5 (ref. 3) and many other regulator proteins⁴. Embedded in the 709-amino-acid polypeptide chain of SWI5 are three tandemly repeated zinc-finger motifs. Because the zinc fingers of TFIIIA are known to bind to DNA, it is probable that in the case of SWI5 these finger motifs also play an important, but not necessarily exclusive, role in the sequence-specific binding of the protein to DNA. To test this prediction we have expressed the 89-amino-acid sequence of the domain containing the three zinc fingers of SWI5 in *Escherichia coli* as a cleavable fusion protein, purified under denaturing conditions and folded *in vitro*. This experimental approach allows us to study directly both the metal requirement and DNA-binding properties of the isolated polypeptide. We find that zinc is required for specific DNA recognition and, most significantly, DNaseI protection studies show that the isolated three-fingered domain is sufficient for sequence-specific binding to DNA.

The SWI5 gene determines the mother-cell-specific transcription of the *HO* endonuclease gene that is responsible for the initiation of mating-type switching in yeast *Saccharomyces cerevisiae*^{5,6}. Recent studies of the SWI5 gene and its product show that it encodes a protein which recognizes a specific sequence(s) located in the promoter of the *HO* gene³. Of the three tandem zinc-finger motifs found in a positively charged region located near the carboxy terminus of the 709-residue polypeptide chain of SWI5 (ref. 3), the first two finger sequences are particularly similar to those of TFIIIA from *Xenopus laevis*, in that they contain the characteristic pairs of Cys and His as well as the hydrophobic Phe, Tyr and Leu at conserved positions (Fig. 1b). For TFIIIA there is now good evidence that each of the 30-amino-acid repeats forms an independently folded mini-domain in which zinc is tetrahedrally coordinated to the two invariant pairs of Cys and His^{1,7-9}. The similarity between the repeated sequence units of SWI5 and those of TFIIIA suggests

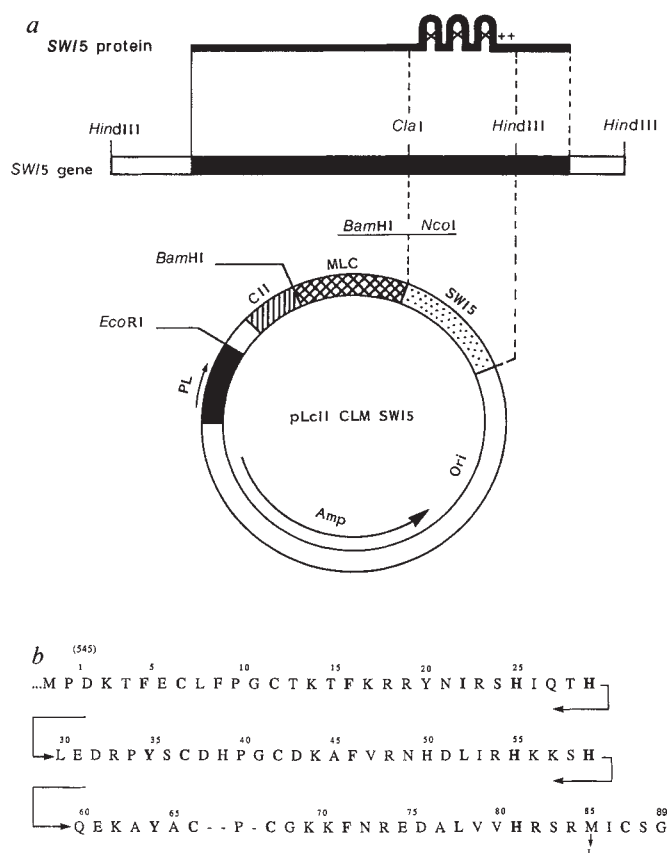


Fig. 1 Expression in *E. coli* of the SWI5 polypeptide containing three zinc-finger motifs. **a**, Construction of the cleavable fusion-protein expression vector pLcII CLM SWI5. The fusion protein consists of the 31 amino-terminal residues of the λ cII protein, 100 amino-acid residues encoded by the myosin light-chain complementary DNA and the SWI5 three zinc-finger motifs. **b**, Sequence of the 89-residue polypeptide containing three zinc-finger motifs in tandem. Pro 1 is amino-acid 545 in the sequence of the SWI5 protein; bold letters, residues conserved between SWI5 finger motifs and those of TFIIIA; dash, gaps in the alignment.

Methods. The ClaI-HindIII fragment of the SWI5 gene³ containing the zinc-finger motifs was cloned into the BamHI-HindIII site of the fusion protein expression vector pLcII via NcoI linker¹⁰. The EcoRI-HindIII fragment of this vector was then cloned into the EcoRI-HindIII site of M13mp19. To produce a polypeptide consisting of the three zinc-finger motifs, a stop codon was introduced at Gly 90. The EcoRI-HindIII fragment of the mutated SWI5 M13 DNA was cloned back into pLcII to form pLcII SWI5. A high yield of fusion proteins was obtained only after insertion in the reverse orientation of a BamHI fragment of myosin light-chain cDNA (CLM) into the BamHI site of the above vectors to create expression vectors pLcII CLM SWI5.

that in SWI5 these motifs also form zinc-containing structural mini-domains that are involved in binding to DNA.

Are the three zinc-finger motifs of SWI5 alone sufficient for sequence-specific DNA binding? Is zinc a necessary component of the folded polypeptide and what sequence(s) of the *HO* promoter does the isolated three-fingered domain recognize? To answer these questions we have expressed the putative DNA-binding domain of SWI5 in *E. coli*. Figure 1a shows the cloning strategy used. The 89-residue polypeptide chain was expressed as part of a cleavable fusion protein using the pLc II expression vector¹⁰. A fragment of the chicken myosin light chain was cloned in the reverse orientation to reduce the overall positive charge of the fusion protein and so make the inclusion bodies that are formed in *E. coli* soluble in urea. The SWI5 polypeptide was liberated from the fusion protein by cyanogen-bromide

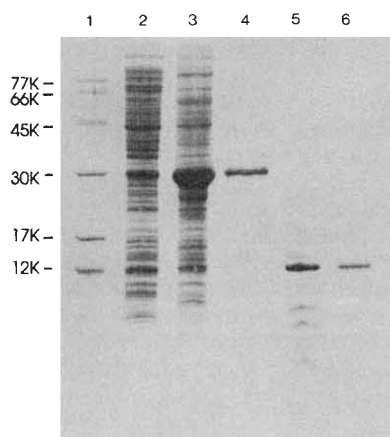


Fig. 2 Expression and purification of the polypeptide containing the three zinc-finger motifs of SWI5. Protein samples at various stages of purification were analysed in a 18% SDS polyacrylamide gel and stained with coomassie blue. Lanes: 1, relative-molecular-mass marker; 2, total cellular extract; 3, inclusion bodies; 4, fusion protein purified on a CM-Sepharose column; 5, purified fusion protein cleaved with CNBr; 6, the cleaved SWI5 finger domain after purification on a CM-Sepharose column.

Methods. Protein synthesis was carried out by heat induction of QY13 strains harbouring pLcHICLSWI5 and inclusion bodies were prepared as previously described¹⁰. The inclusion body fraction from 40-l culture was dissolved in 150 ml 8 M urea, 25 mM Tris phosphate pH 7.5, 1 mM EDTA, 1 mM dithiothreitol (DTT) and applied to a 4 × 10 cm CM-Sepharose-CL (Pharmacia) column equilibrated with the same buffer. The fusion protein was then eluted with an NaCl gradient. The purified fusion protein was cleaved with CNBr and the three-fingered polypeptide isolated on the same CM-Sepharose-CL column. From this expression and purification procedure we obtain 5–10 mg of liberated SWI5 peptide per litre of *E. coli* culture. Folding of the finger domain was carried out by dialysis, first against 4 M urea, 1 mM Tris-acetate pH 5.0, 1 mM DTT, and then against 20 mM Tris-chloride pH 7.5, 70 mM KCl, 2 mM MgCl₂, 200 μM DTT, 100 μM zinc acetate.

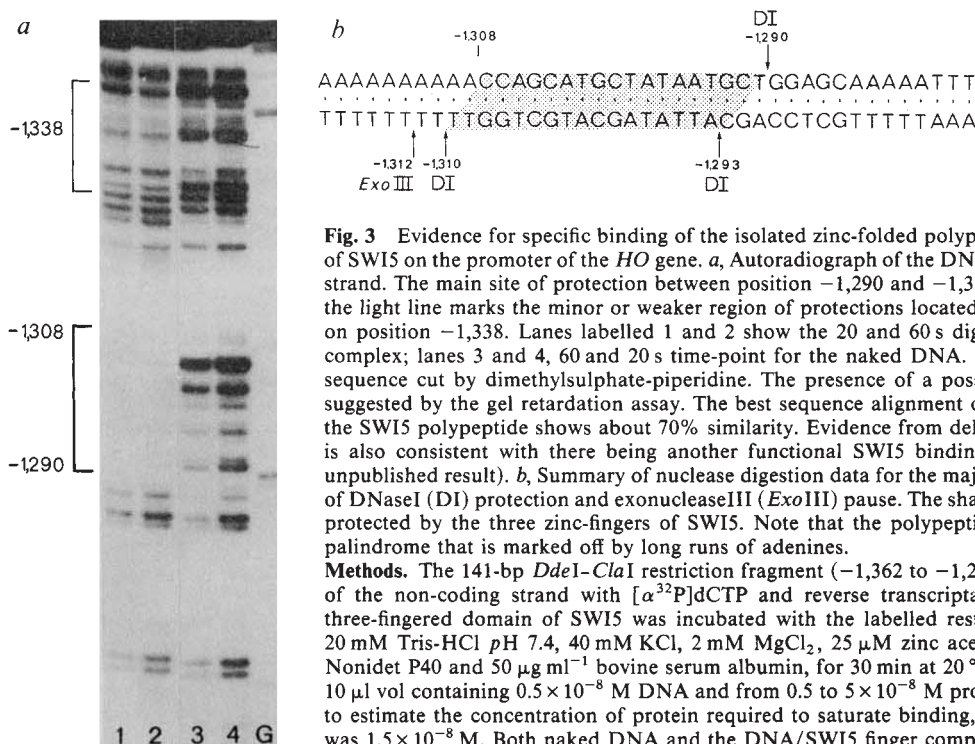


Fig. 3 Evidence for specific binding of the isolated zinc-folded polypeptide containing the three-finger motifs of SWI5 on the promoter of the *HO* gene. *a*, Autoradiograph of the DNaseI protection pattern of the non-coding strand. The main site of protection between position –1,290 and –1,308 is marked on the left by a heavy line; the light line marks the minor or weaker region of protections located upstream of the major one and centred on position –1,338. Lanes labelled 1 and 2 show the 20 and 60 s digestion time-points of the protein/DNA complex; lanes 3 and 4, 60 and 20 s time-point for the naked DNA. Lane G, the position of guanines in the sequence cut by dimethylsulphate-piperidine. The presence of a possible second binding site had also been suggested by the gel retardation assay. The best sequence alignment of the two regions of DNA protected by the SWI5 polypeptide shows about 70% similarity. Evidence from deletion analysis of the *HO* gene promoter is also consistent with there being another functional SWI5 binding site upstream of the first one (K.N., unpublished result). *b*, Summary of nuclease digestion data for the major binding site. Arrows mark the borders of DNaseI (DI) protection and exonucleaseIII (ExoIII) pause. The shaded area indicates the DNA sequence protected by the three zinc-fingers of SWI5. Note that the polypeptide binds to the 5' side of an imperfect palindrome that is marked off by long runs of adenines.

Methods. The 141-bp *DdeI*-*ClaI* restriction fragment (–1,362 to –1,221) was labelled selectively at the 3' end of the non-coding strand with [α^{32} P]dCTP and reverse transcriptase. After purification and folding, the three-fingered domain of SWI5 was incubated with the labelled restriction fragment in a buffer containing 20 mM Tris-HCl pH 7.4, 40 mM KCl, 2 mM MgCl₂, 25 μM zinc acetate, 0.5 mM DTT, 3% glycerol, 0.05% Nonidet P40 and 50 μg ml^{–1} bovine serum albumin, for 30 min at 20 °C. Binding reactions were carried out in 10 μl vol containing 0.5 × 10^{–8} M DNA and from 0.5 to 5 × 10^{–8} M protein. The gel retardation assay was used to estimate the concentration of protein required to saturate binding, which for the experiment shown above was 1.5 × 10^{–8} M. Both naked DNA and the DNA/SWI5 finger complex were digested at 20 °C and at a final DNaseI concentration of 1 μg ml^{–1}, for increasing lengths of time (15–120 s) to obtain similar extents of digestion of the two samples. The fragments produced were analysed in an 8% sequencing gel containing 8 M urea.

cleavage at the Met residue situated immediately before the first finger motif (Fig. 1b). But to obtain the three finger motifs intact, the naturally occurring Met 85 in the third finger was replaced by Leu (Fig. 1b). This mutation did not affect the DNA-binding specificity of the folded fusion protein, as judged by DNaseI protection experiments. The 89-residue polypeptide was purified to homogeneity on a CM-Sepharose column in the presence of 8 M urea (Fig. 2) and then folded or renatured by dialysis under reducing conditions in the presence and absence of zinc ions. Its binding activity was analysed by the gel retardation assay (not shown) and DNaseI protection studies (see below). For the polypeptide folded in the presence of zinc a two to three molar excess of protein was required to saturate binding, indicating that 30–50% of the molecules were correctly folded. For the same batch of polypeptide folded in the absence of zinc, no binding was observed at the concentration of protein that saturates binding for the zinc-folded polypeptide. This supports the proposal that zinc is important for maintaining the tertiary

structure required for sequence-specific recognition, and is in agreement with earlier studies on TFIIIA^{7,11}. Frankel *et al.*⁹ have expressed one of the finger motifs of TFIIIA in *E. coli* but their zinc-folded finger domain does not show sequence-specific binding to the DNA, and therefore there is no direct evidence for correct folding.

The location and size of the binding site of the isolated and folded three-fingered polypeptide of SWI5 was identified from DNaseI protection studies. A restriction fragment known to contain the SWI5 binding-site, as identified using yeast extracts³, was end-labelled and lightly digested in the presence and absence of protein. The autoradiograph in Fig. 3a shows that the SWI5 polypeptide protects from DNaseI cleavage two sites on the DNA, a major and a minor one. The minor site is discussed in the legend to Fig. 3. The major, or most strongly protected site extends between positions –1,290 and –1,308 of the non-coding strand (Fig. 3a) and between positions –1,293 and –1,310 of the coding strand (data not shown). These two

sets of data define the 3' border of protection for both strands to base pair (bp) -1,291. The 5' border of DNaseI protection is somewhat less certain because the cleavage rate of DNaseI is very slow within the long runs of adenines found there. Additional information about the 5' border was obtained from a digestion study using exonuclease III that gives a pause on the coding strand at position -1,312 (data not shown). The summary of the data in Fig. 3b reveals that the SWI5 polypeptide protects about 20 bp of DNA from nuclease cleavage. When the large diameter of the nucleases used is taken into consideration, the length of DNA bound by the three tandemly repeated zinc-finger motifs of SWI5 is probably closer to 15 bp. Assuming that one molecule of protein is bound, the size of the binding site is in this case consistent with our earlier proposal that each finger domain binds, on the average, about half a double-helical turn of DNA¹².

The main binding site of the isolated three-fingered polypeptide of SWI5 is in the region of the *HO* promoter required for the SWI5 dependent transcriptional activation *in vivo*³. It also falls in the region protected against DNaseI digestion by yeast extract in which the entire SWI5 protein had been over-expressed³. The 3' border of protection by the three-fingered polypeptide and by SWI5 coincide, whereas the 5' border in the latter case appears to extend across the run of adenines (Fig. 3b).

This difference might be explained by the fact that SWI5 is a large protein that could protect a longer stretch of DNA from DNaseI.

Here we have shown that three zinc fingers of SWI5 alone are sufficient for directing binding of the transcription factor to its recognition sequence in the promoter region of the *HO* gene. Although gene-deletion analysis of the yeast regulatory protein ADR1 has implicated a region containing homologous finger motifs in DNA binding¹³, our study directly demonstrates the role of zinc-finger motifs in sequence-specific binding to DNA. Our results support the idea that the finger motifs form zinc-binding mini domains that together constitute the DNA-binding domain of the larger transcription factor. This DNA-binding domain appears to be both structurally and functionally independent of the rest of the protein.

We thank C. Young for cell culture, L. Fairall for help with exonuclease III digestion studies, T. Creighton and A. Klug for discussions and A. Klug and H. Nelson for comments on the manuscript. Y.N. was supported by the TOYOBO Biotechnology Foundation.

Note added in proof: After this paper was submitted, Kadonaga *et al.* (*Cell* **51**, 1079-1090 (1987)) reported that the DNA binding activity of transcription factor SP1 is located in a 168 amino acid sequence that contains three zinc-finger motifs.

Received 22 November 1987; accepted 4 February 1988.

1. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1604-1614 (1985).
2. Brown, R. S., Sander, C. & Argos, P. *FEBS Lett.* **186**, 271-274 (1985).
3. Stillman, D. J., Bankier, A., Seddon, A., Gorehout, G. & Nasmyth, K. A. *EMBO J.* **2**, 485-494 (1988).
4. Klug, A. & Rhodes, D. *Trends biochem. Sci.* **12**, 464-469 (1987).
5. Stern, M., Jensen, R. & Herskowitz, I. *J. molec. Biol.* **178**, 853-868 (1984).
6. Nasmyth, K., Seddon, A. & Ammerer, G. *Cell* **49**, 549-558 (1987).

7. Diakun, G. P., Fairall, L. & Klug, A. *Nature* **324**, 689-691 (1986).
8. Tsai, T. Y., Van Den Berg, D. J. & Korn, L. J. *Nucleic Acids Res.* **14**, 2187-2200 (1986).
9. Frankel, A. D., Berg, J. M. & Pabo, C. O. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4841-4845 (1987).
10. Nagai, K. & Thögersen, H. C. *Meth. Enzym.* **143**, 461-481 (1987).
11. Hanas, J. S., Hazuda, D. J., Bogenhagen, D. F., Wu, F. Y. H. & Wu, C. W. *J. biol. Chem.* **258**, 14120-14125 (1983).
12. Rhodes, D. & Klug, A. *Cell* **46**, 123-132 (1986).
13. Hartshorne, T. A., Blumberg, H. & Young, E. T. *Nature* **320**, 283-287 (1986).

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article
Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem

☐ Other (please specify) _____

☐ I am interested in sending my order by mail.

☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

The iscom antigen-presenting system

B. Morein

The immunostimulating complex, or iscom, provides an effective means of presenting antigens to the immune system. Vaccines for influenza, hepatitis B and AIDS are in the offing.

WE can soon celebrate the bicentennial anniversary of the first vaccine — the vaccine against smallpox. Although it has taken nearly 200 years to wipe out smallpox as a disease, its eradication constitutes a great achievement for vaccine technology. Vaccines against polio, measles and foot-and-mouth disease now have been developed according to the same traditional concept: inoculation using whole dead or live attenuated microorganisms. These traditional vaccines are effective against microorganisms which cause acute infections leading to death or complete recovery, but vaccination attempts have failed against microorganisms which cause chronic or slow infections. New approaches are needed for formulating vaccines against these pathogens, which persist in the host and evade the immune system by hiding a potentially immunogenic antigen or antigenic determinant. A vaccine against such a virulence mechanism must circumvent this strategy; isolating and reorganizing the potentially immunogenic antigens offers one approach.

New approaches to vaccine formulation are also required in order to benefit from the new methods of producing defined antigens by gene technology or by the chemical synthesis of antigenic epitopes. Such antigens are in general poorly immunogenic, unless they are organized in multimeric copies as defined particles such as protein micelles or virosomes. But micelles have limited immunogenicity, as exemplified by the fact that a well-defined experimental micelle vaccine containing the envelope proteins of parainfluenza-3 virus (PIV-3) did not induce protective immunity in lambs to challenge with a PIV-3 strain causing pneumonia unless an oil adjuvant was included¹. However, adjuvants frequently are only effective when used in such high doses that they also cause undesirable or pathologic side-effects.

In an attempt to combine a stable and well-defined antigen complex with a built-in adjuvant, the immunostimulating

complex (iscom) was formulated². The adjuvant — Quil A — was included as a matrix. Quil A is an extract from the bark of the *Quillaja saponaria molina* tree found in the Andes between Chile and Argentina. The iscom has an average diameter of 35 nm and is composed of subunits of about 12 nm in size (Fig. 1). The iscom is held together by hydrophobic interactions between the matrix (Quil A), lipids and the amphipathic antigens^{2,3}.

The immunogenicity of antigens presented in iscoms was recently reviewed⁴. Generally, a ten-fold higher immune response is obtained with iscom-bound antigen compared to the same antigen in a

conventional types of vaccines. It is well documented that a whole killed measles virus vaccine does not induce the anti-fusion activity necessary for complete protection, so this distinction is important. In contrast, iscoms containing the fusion and haemagglutinin proteins from the measles virus induced antifusion antibody². Conventional experimental vaccines of whole killed feline leukaemia virus (FeLV) do not induce neutralizing antibody or protective immunity in cats. It is out of the question for FeLV to be used as a live vaccine because of the inherent risk that its genome will be integrated into the host cell genome.

Table 1 Protective immunity induced by iscoms containing various microbial antigens

Antigen	Animal	Disease
Haemagglutinin, neuraminidase (H1N1), influenza virus ¹²	mice	pneumonia
Haemagglutinin measles ¹³	mice	encephalitis
Fusion protein measles ^{13,14}	mice	encephalitis
Toxoplasma surface antigens ¹⁵	mice	lethal
Bovine virus diarrhoea virus envelope protein ¹⁶	sheep	abortion
Feline leukaemia virus, gp70 ⁸	cat	viraemia
Epstein-Barr virus ¹⁷	tamarin monkey	tumour

micelle or *in situ* in a virus particle. Antigens from more than twenty viruses, and from bacteria, parasites and animal cells have been incorporated into iscoms. Serum antibody responses in mice have been studied with several iscom preparations, such as those made from the envelope proteins of influenza virus (H1N1). After both local (intranasal) and subcutaneous applications, classical antibody responses have been obtained. Initially, only IgM antibody is present, but subsequently, antibody from all isotypes of IgG appears⁵.

Potent cell-mediated immune responses have been recorded following immunization of monkeys with iscoms containing cytomegalovirus antigens, measured by the lymphocyte stimulation test⁶. Additionally, two intranasal immunizations with iscoms containing the envelope proteins of influenza virus (H1N1) induced a cytotoxic memory T-cell response⁷. These iscoms were also able to induce a comparatively high secondary cytotoxic T-cell response.

Using iscom-bound antigens, protective immune responses have been induced to a variety of microorganisms in different animal models (Table 1) which deviate positively from responses elicited from

However, iscoms containing gp70 from FeLV induce both neutralizing serum antibody and protective immunity in cats⁸. Antigens produced by recombinant DNA techniques, such as the hepatitis B surface antigen produced in yeast⁹, and various HIV antigens produced in *Escherichia coli* or in animal cells, become highly immunogenic when included into iscoms and may be an effective and safe alternative to vaccination with live virus. Neutralizing antibodies to HIV also have been introduced by some of these preparations¹⁰.

Apart from the difficulty encountered in creating synthetic polypeptides which mimic correctly the antigenic determinant that they represent, there are two major problems with making them immunogenic. First, epitopes of synthetic polypeptides need to be presented in an appropriate antigenic context to antigen-responsive cells of the immune system. Second, the epitopes need to be formulated into a defined structural complex such as that of a micelle or iscom. The work of Lövgren *et al.*¹¹ showed that a small molecule like biotin could be highly immunogenic when a sufficient number of molecules (about 10) were coupled to the haemagglutinin of influenza virus and integrated into iscoms or micelles. The



Fig. 1 Iscoms. Magnification, $\times 125,000$.

biotin molecule in the iscom was considerably more immunogenic than the biotin coupled to micelles or to bovine serum albumin emulsified in Freund's complete adjuvant. Work in my laboratory has confirmed that oligopeptides such as human chorionic gonadotropin conjugated to the haemagglutinin iscom are highly immunogenic provided a sufficiently high number of peptides are coupled per carrier molecule.

The amount of Quil A integrated into an iscom is 5 to 10 per cent of the antigen dose. In general, such iscoms are innocuous, but depending on the different preparation methods used⁴, a separation of the iscom from the free Quil A may be required to reduce the amount of Quil A in a potential vaccine. A mouse weighing 20 g, for example, can tolerate a dose of 10 µg of Quil A with basically no toxic reactions. Free Quil A does not contribute to the immunogenicity of iscoms, but high amounts may induce side effects in mice.

The iscom is stable, amenable to lyophilization, and can be produced on a large scale. These properties, coupled with its efficiency in presenting antigens to the immune system, indicate that the iscom should be useful in the production of vaccines. The first commercial iscom vaccine, against influenza in horses, has been licensed in Sweden, and iscom technology will be applied in the development of human vaccines when toxicology studies are completed. Preliminary studies performed with an experimental measles vaccine at the National Institute of Public Health and Environmental Hygiene at Bilthoven in the Netherlands look promising. The ability of iscoms to be used for local immunization, in contrast to earlier non-replicatory vaccines or antigen preparations, will make them even more useful. □

Bror Morein is at the Department of Virology, National Veterinary Institute, Biomedical Center, Box 585, S-751 23 Uppsala, Sweden. For more information fill in reader service number 100.

- Morein, B. & Simmons, K. *Vaccine* 3, 83-93 (1985).
- Morein, B., Sundquist, B., Höglund, S., Dalsgaard, K. & Osterhaus, A. *Nature* 308, 457-460 (1984).
- Lövgren, K. & Morein, B. *Biotech. appl. Biochem.* (in the press).
- Morein, B., Lövgren, K., Höglund, S. & Sundquist, B. *Immunol. Today* 8, 333-338 (1987).
- Lövgren, K. *Scand. J. Immunol.* 2, 241-245 (1988).
- Wahren, B., Nordlund, S., Åkeson, A., Sundquist, V.-A. & Morein, B. *Med. microbiol. Immunol.* 176, 13-19 (1987).
- Jones, P. D., Tha Hla, R., Morein, B., Lövgren, K. & Ada, G. L. (in preparation).
- Osterhaus, A. *et al. J. Immunol.* 135, 591-596 (1985).
- Howard, C. R., Sundquist, B., Allen, J., Brown, S. E., Chen, S.-H. & Morein, B. *J. Virol.* 68, 2281-2289 (1987).
- Åkerblom, L., Villacres, M. & Morein, B. (in preparation).
- Lövgren, K., Lindmark, J., Pipkorn, R. & Morein, B. *J. Immunol. Meth.* 98, 137-143 (1987).
- Sundquist, B., Lövgren, K. & Morein, B. *Vaccine* 6, 49-53 (1988).
- Varsanyi, T., Morein, B., Löve, A. & Norrby, E. *J. Virol.* 61, 3896-3901 (1987).
- de Vries, P. *et al. J. gen. Virol.* (in the press).
- Uggla, A. *et al.* (in preparation).
- Ahlénius, S., Carlsson, U. & Sundquist, B. *Proc. 14th Wld Congr. Dis. Cattle* 1, 1326-1327 (1986).
- Morgan, A. J., Finerty, S., Lövgren, K., Scullion, F. T. & Morein, B. (in preparation).

MAbs on the market

An increasingly diverse selection of monoclonal antibodies are available. Some of the newest — plus other immunology products — are described below.

AN immunoassay for the measurement of **auto-antibodies to double-stranded DNA** using only 5 µl of human serum is available from BradSure Biologicals (*Reader Service No. 101*). The Progen enzyme immunoassay allows quantitative serological determination of dsDNA auto-antibodies through comparison with a human standard graph of four titration points, adjusted to the World Health Organization standard. BradSure says the coated dsDNA, isolated from calf thymus, is devoid of single-stranded regions. The £180 (UK) Progen assay kit contains enough reagent for 60-90 simple determinations.

For the detection of B-cell surface antigens in formalin-fixed, paraffin-embedded sections, ICN ImmunoBiologicals offers its **lymphoma identification kit** (*Reader Service No. 102*). The \$275 (US) kit contains a panel of three monoclonal antibodies: LN-1, LN-2 and LN-3.



Lymphoma ID kits from ICN.

LN-1 reacts with the surface membrane and cytoplasm of germinal B-cells, LN-2 reacts with the nuclear membrane and cytoplasm of mantle zone and germinal centre B-cells and interdigitating histiocytes, and LN-3 reacts with HLA-DR antigen. ICN says the panel of antibodies links the ability to assess the immunologic phenotype of neoplastic lymphocytes with morphologic analysis, properties useful for phenotyping malignant lymphomas.

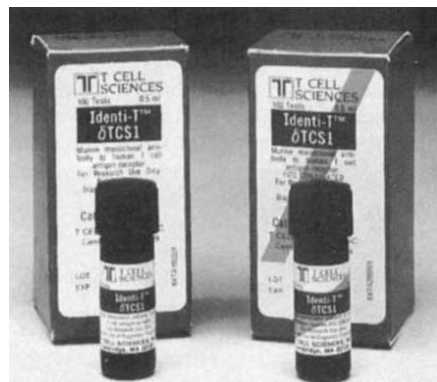
Software

AssayZap is the name of a universal **assay calculator** for Apple Macintosh computers from Biosoft (*Reader Service No. 103*). Biosoft says AssayZap offers the familiar regression fit to the logistic equation for calculating RIAs, ELISAs and IRMAs, plus an interactive visual curve fitting technique that allows all standard curves to be fitted, whatever their shape. Least squares minimization routines for

two-parameter log-logit fitting, and unweighted and weighted four-parameter fit are included. The program maintains a record of all previous assays processed, and permits the comparison of stored curves with the current standard curve. AssayZap can handle assays of up to 2,000 samples, and each assay may include up to four standard curves. Biosoft sells the AssayZap program for \$249 (US). A demonstration version and documentation can be obtained for a refundable charge of \$20 (US).

The ImmunoFit software from Beckman is an **EIA/RIA data analysis package** designed for use with an IBM PC and a Beckman liquid scintillation counter or the Beckman Biomek 1000 automated laboratory workstation (*Reader Service No. 104*). More than fifty combinations of axes, selections and curve fits are available with ImmunoFit, and each can be manipulated using the mouse. Data is graphically displayed on screen so that the best choice for X and Y axis and curve fit may be made using the statistics provided. ImmunoFit produces a complete report including a standards table, standard curve plot, quality control parameters, controls, unknowns and a precision profile plot, that can be printed out or stored in memory. Beckman's ImmunoFit costs £1,200 (UK).

For studying T lymphocytes which express the $\gamma\delta$ **T-cell antigen receptor**, T Cell Sciences has a monoclonal antibody to the δ chain — **Identi-T δ TCS1** (*Reader Service No. 105*). Identi-T δ TCS1 can be used in flow cytometry, immunochemical staining and immunoprecipitation procedures, and its mitogenic properties can be used to advantage in the growth and cloning of $\gamma\delta$ T-cells. Identi-T δ TCS1 is sold in a liquid,



For the δ chain of the T-cell antigen receptor.

ready-to-use form as both purified and fluorescein isothiocyanate (FITC) conjugated antibody, for \$355 and \$395 (US), respectively, for 100 tests.

Bethesda Research Laboratories has added **affinity-isolated secondary antibodies** to its line of molecular biology and immunology products (*Reader Service No. 106*). All secondary antibodies from BRL are purified by affinity isolation to reduce nonspecific background and increase sensitivity, and come with a satisfaction guarantee. The range of antibodies offered includes goat anti-mouse IgG (H and L chain specific) and subclasses IgG₁, IgG_{2a}, IgG_{2b}, and IgG₃; goat anti-mouse IgA, IgM, kappa chain, lambda chain, and a combination of IgG, IgA and IgM; goat anti-rabbit IgG (H and L chain specific); goat anti-human IgG (F_c); goat anti-human IgM; and goat anti-rat IgG (H and L chain specific). Antibodies conjugated to horseradish peroxidase, alkaline phosphatase and FITC are also available.

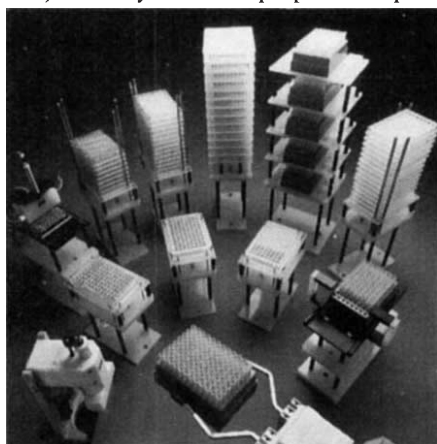
Amersham now has a range of **receptor monoclonal antibodies** for the localization of cellular receptors (*Reader Service No. 107*). Antibodies in the range include those for the receptors of epidermal growth factor, transferrin, interleukin-2 and low density lipoprotein. They may be used on frozen sections and on cell cultures or in immunoprecipitation studies, in conjunction with biotinylated second antibodies, streptavidin-enzyme or streptavidin-gold systems, or with Amersham's DAB silver enhancement kit.

MabLab is the name of the production-scale **automated antibody purification system** from Oros (*Reader Service No. 108*). The MabLab system consists of a computer module, a separation module based on a protein A column, and a staging module. The software is an expert system control program that interprets information received from sensing devices and automatically controls system performance. The \$100,000 (US) MabLab produces up to 2 g per day of almost all IgG

and some IgM. Oros says the MabLab can work out how best to purify a totally uncharacterized monoclonal antibody and achieve purity levels of over 95 per cent within hours. An automatic cleaning cycle between batches ensures no cross-contamination between products.

Microplates and more

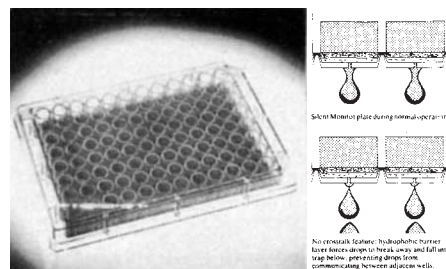
Zymark brings **robotics** to immunology with its Microplate Management System, capable of running fully automated ELISA procedures (*Reader Service No. 109*). The system can prepare samples,



The array of microplate modules from Zymark.

coat microplates, perform ELISAs and feed the microplates into a reader. The user can configure the system to execute many combinations of incubation times and wash cycles; tight tolerance on the incubation times makes quenching the reaction unnecessary. Zymark says throughput with the system ranges from 35 to 50 plates daily. The Microplate Management System includes a robot and microplate pipetting and stacking modules, and ranges in price from \$45,000 to \$55,000 (US).

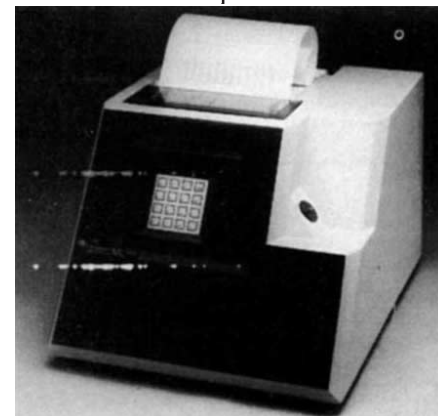
The Silent Monitor **membrane-bottomed 96-well test plate** from Pall has a three-layer well bottom composition that Pall says eliminates cross-talk between wells (*Reader Service No. 110*). The first layer of Pall's well-bottoms is an intrinsically hydrophobic nylon 66 membrane, which



No cross-talk between wells is the Silent Monitor microplate's claim to fame.

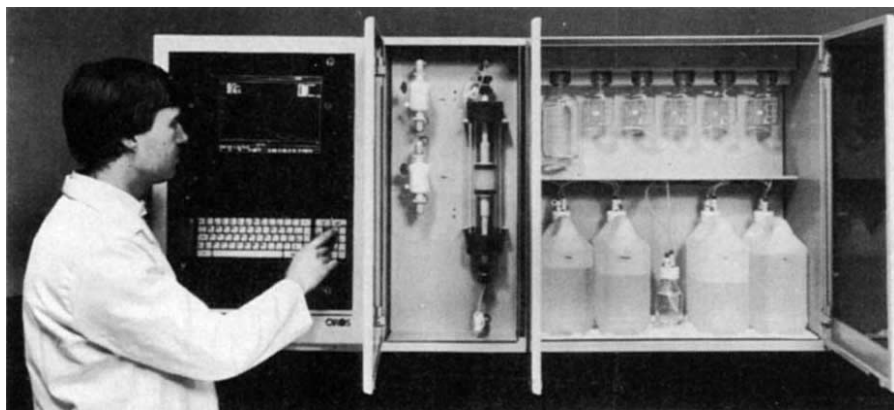
also serves as a solid phase filter separation barrier. The second layer, made of a moderately hydrophobic open polymer fibre, ensures even drainage and forms a non-wettable barrier at the perimeter of each well. The third layer's strongly hydrophobic fluoropolymers direct flow under a vacuum through a small control orifice to encourage the separation of drops and prevent liquid spreading from well to well. The Silent Monitor plates are available with Immunodyne activated membrane for covalent protein binding, Biotodyne A & B amphoteric or positively charged membranes, and Loprodyne low protein-binding membranes. The plates cost £50 (UK) for a pack of five.

Dynatech's automated **microplate scanning luminometer** is designed for chemi- and bio-luminescence assays (*Reader Service No. 111*). The model ML1000 Microlite plate scanner can be



Chemi- and bioluminescence assays get automated with Dynatech's device.

programmed to perform a variety of operations, with single and multiple scan options. The user specifies the numbers of test samples, dispense volume and gain factor, and selects between six alternative operation modes and data presentation modes. Dual and integrate time with peak, half and full integral values are presented graphically by the built-in printer. The Microlite can handle both traditional 96-well plates and removable-strip or removable-well composite plate systems. The £13,500 (UK) Microlite comes with a precision volumetric dispensing unit for applications requiring reagent injection, and an RS232 port for transmitting data to a computer.



The MabLab from Oros produces up to 2 g per day of antibody practically independently.

Immuno-reagents

Sera-Lab specializes in serum and antibodies, and their 1988 **catalogue** reflects the breadth and depth of their offerings (*Reader Service No. 112*). An additional 100 monoclonal antibodies have been added to Sera-Lab's range, bringing the total of monoclonal antibodies they offer to human, mouse, rat, bacterial and viral determinants to 376. A comprehensive range of polyclonal antibodies and second step reagents, and streptavidin-biotin range conjugates and antibodies are also new to the 1988 edition.

Three new reagents for **cytokine research** are now available from Genzyme Corporation: recombinant human granulocyte colony stimulating factor, recombinant human macrophage colony stimulating factor and monoclonal anti-human granulocyte-macrophage colony stimulating factor (*Reader Service No. 113*). The G-CSF and M-CSF are purified by ion-exchange chromatography and reverse-

phase HPLC, and are supplied as 5,000 CFU per vial with an activity of greater than 1×10^6 per mg of protein, says Genzyme. Each vial costs \$125 (US). The G-CSF is produced in yeast using a gene from the human 5637 bladder carcinoma cell line, and the M-CSF is produced in yeast using a truncated version of the gene isolated from the MIA-PaCa-2 human pancreatic tumour cell line. The monoclonal anti-GM-CSF is a 95 per cent pure IgG, murine monoclonal raised against purified recombinant human GM-CSF. Genzyme supplies the anti-GM-CSF in 0.5 mg quantities for \$250 (US).

The new catalogue of **immunology research reagents** from Organon Teknika contains a broad range of animal research reagents (*Reader Service No. 114*). Included are diagnostic kits for the identification of viral antibodies in laboratory animals, antisera to various animal proteins provided in conjugated and unconjugated form, fluorescein and protein A. A variety of monoclonal antibodies and antisera for forensic medicine and a range of bovine plasma products are also listed.

Cellon Sarl has just introduced a number of reagents for the detection and diagnosis of **autoimmune disease** to its line of immunology/immunodiagnostics (*Reader Service No. 115*). Immunoaffinity purified antigens include SS-B (La), Sm, Sm/RNP complex and rabbit or calf thymus nuclear extracts containing the RNP, Sm, SS-A, SS-B, JO-1, PM₁ and Scl-70 (Topoisomerase 1) determinants. These are suitable for ELISA, immunodiffusion or other immunoassays for the study of systemic lupus erythematosus, mixed connective tissue disease, scleroderma sjögrens syndrome, polymyositis or dermatomyositis. Autoimmune control sera screened for both hepatitis B surface antigen and antibody to HIV are also available for anti-Sm, RNP, SS-A (Ro), SS-B (La), Scl 70 (Topoisomerase 1), mitochondrial and thyroid (microsomal) autoantibodies. Suggested protocols are included for assay development.

Cambio is offering reduced prices on **interleukins** produced by Janssen Biochemica (*Reader Service No. 116*). Bulk human interleukin-2 in both crude and PHA-free forms will be available for £385 and £580 (UK), respectively, for a 500 ml bottle until the end of March. Cambio also sells Janssen's hybridoma growth factor in purified and crude preparations, and a wide range of antibodies with subclass typing antisera.

Serum and substitutes

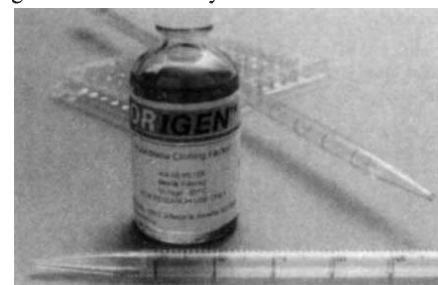
Life Science Laboratories Ltd has a new **serum substitute** that it says eliminates the variability encountered with natural serum, decreases the total protein content



Serum substitutes from Life Science Labs.

in the medium, and eliminates the presence of immunoglobulins, simplifying product purification (*Reader Service No. 117*). The Ultrosor HY medium was specially developed for the growth of hybridomas and lymphoblastoid cells. Life Science Laboratories reports that basal medium supplemented with only 2 per cent of the substitute can produce growth rates at least equal to those achieved with 10–20 per cent natural sera. The Ultrosor HY serum costs £187.50 (UK) for a package of 20 vials; each makes 10ml.

A **hybridoma cloning factor** called Origen from Igen, Inc. enhances the growth of murine hybridoma cell lines cul-



Igen's hybridoma cloning factor.

tured at limited cell densities (*Reader Service No. 118*). Igen says Origen is more efficient in fusion and hybridoma cloning than feeder layers and conditioned media alternatives, because it is a partially purified filter-sterilized material derived from continuous macrophage cell lines. Origen comes in 50 ml bottles at 10× concentration for \$150 (US), and is stable for a minimum of six months when stored at –20 °C according to the company.

Biocell's 48-page **product catalogue** has three useful reference charts describing the properties of human diagnostic grade serum, human and animal tissue culture serum, and albumins (*Reader Service No. 119*). Over 200 products for diagnostics and research are listed in the catalogue, including immunoassay grade materials.

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE – HOMOGENEOUS –
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5–10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIABORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.49



Cell Injection?

Singer MK III
Singer Microforge

Singer Instrument Co. Ltd.

Treborough Lodge
Roadwater, Watchet
Somerset, England, TA23 0QL
Tel: Washford (0984) 40226
Telex: 337492 Comcab G

Reader Service No.50